

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2000, 18:11:54 ; Search time 1073.96 Seconds
(without alignments)
1073.635 Million cell updates/sec

Title: US-09-300-482-569

Perfect score: 264
Sequence: 1 ctgacttgagcaattccca.....gacatgctcatgatcttta 264

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_cm:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: gb_htg1:*
31: gb_htg2:*
32: gb_in1:*
33: gb_in2:*
34: em_ba1:*
35: em_ba2:*
36: em_hum3:*
37: em_hum4:*
38: gb_pr4:*
39: gb_htg3:*
40: gb_htg4:*
41: gb_htg5:*
42: gb_htg6:*
43: gb_htg7:*

44: em_htg1:*
45: em_htg2:*
46: em_htg3:*
47: em_hum5:*
48: gb_pl3:*
49: gb_pr5:*
50: gb_htg8:*
51: gb_htg9:*
52: gb_htg10:*
53: gb_htg11:*
54: gb_htg12:*
55: gb_htg13:*
56: gb_htg14:*
57: gb_in3:*
58: gb_htg15:*
59: gb_htg16:*
60: gb_htg17:*
61: em_htg4:*
62: em_htg5:*
63: em_htg6:*
64: em_htg7:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_vl1:*
73: gb_vl2:*
74: gb_ba3:*
75: em_htg8:*
76: em_htg9:*
77: em_htg10:*
78: em_htg11:*
79: em_htg12:*
80: em_htg13:*
81: em_htg14:*
82: em_htg15:*
83: em_htg16:*
84: em_htg17:*
85: em_htg18:*
86: em_htg19:*
87: em_htg20:*
88: em_htg21:*
89: em_htg22:*
90: em_htg23:*
91: gb_pr6:*
92: gb_pr7:*
93: gb_sts1:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	51.2	19.4	7218	5 166494
2	38.2	14.5	1552	48 ST095923
3	38.2	14.5	147098	39 AC011452
4	38.2	14.5	164313	30 AC010624
5	38.2	14.5	177590	50 AC021154
6	37.6	14.2	167390	10 AC007263
7	37.6	14.2	200542	71 CNS01RG3
8	37.4	14.2	163854	50 AC021590
9	37.2	14.1	79452	42 AC016536
10	37.2	14.1	107109	91 HSDJ781B1
11	36.8	13.9	167079	67 AL160400
12	36.8	13.9	187999	54 AC026928

```

c 13 36.8 13.9 194963 51 AC022699
c 14 36.6 13.9 24721 91 HSL190H5
c 15 36.6 13.9 181517 68 AL353773
c 16 36.6 13.9 223263 67 AL158068
c 17 36.4 13.8 152590 40 AC012027
c 18 36.4 13.8 212382 39 AC010936
c 19 36.2 13.7 62109 38 AL359984
c 20 36 13.6 569 94 HUM05530
c 21 36 13.6 152370 66 AL138894
c 22 36 13.6 170132 71 CNS05TDS
c 23 36 13.6 186841 40 AC011974
c 24 35.6 13.5 189070 7 API48542
c 25 35.4 13.4 118591 43 AC020202
c 26 35.4 13.4 177987 31 AC008948
c 27 35.4 13.4 188942 52 AC026280
c 28 35.4 13.4 260580 32 AE003791
c 29 35.2 13.3 76042 7 AC006921
c 30 35.2 13.3 115725 53 AC025481
c 31 35.2 13.3 158546 51 AC022876
c 32 35.2 13.3 167312 70 AP001024
c 33 35.2 13.3 173081 59 AC068996
c 34 35.2 13.3 178817 68 AL354931
c 35 35.2 13.3 195944 70 AP002353
c 36 35.2 13.3 200038 43 AC018423
c 37 35 13.3 18357 49 HSL142C11
c 38 35 13.3 118357 69 AL391220
c 39 35 13.3 151102 70 AP001019
c 40 35 13.3 159747 70 AP001496
c 41 35 13.3 177055 70 AP001496
c 42 35 13.3 177629 58 AC060829
c 43 35 13.3 178676 59 AC026965
c 44 35 13.3 182945 59 AC069004
c 45 35 13.3 183005 67 AL161732

```

ALIGNMENTS

```

RESULT 1
LOCUS 166494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE
1. (bases 1 to 7218)
Dornier, F., Schefflinger, F. and Falkner, F. Gunter.
Recombinant fowlpox virus
Patent: US 5670367-A 14-23-SEP-1997;
JOURNAL
FEATURES
Location/Qualifiers
source
1..7218
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match 19.4%; Score 51.2; DB 5; Length 7218;
Best Local Similarity 5.5%; Pred. No. 0.00036;
Matches 11; Conservative 128; Mismatches 61; Indels 0; Gaps 0;

```

```

Db 1371 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1430
QY 195 ttacgatacaatgcctcca 214
Db 1431 YYYYYGTACCAAAATCTCTA 1450

```

```

RESULT 2
LOCUS STU95923 1552 bp mRNA PLN 11-MAY-1997
DEFINITION Solanum tuberosum transaldolase (Potrall) mRNA, complete cds.
ACCESSION STU95923
VERSION U95923.1 GI:2078349
KEYWORDS
SOURCE
potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum.
REFERENCE
1. (bases 1 to 1552)
Mehs, C.P., Allen, P.V., Friedman, M. and Belknap, W.R.
Cloning and expression of transaldolase from potato
Plant Mol. Biol. 32 (3), 447-452 (1996)
MEDLINE
97134933
REFERENCE
2. (bases 1 to 1552)
Mehs, C.P., Allen, P.V., Friedman, M. and Belknap, W.R.
Direct Submission
Submitted (31-MAR-1997) C1U, USDA/ARS, 800 Buchanan Street, Albany,
CA 94710, USA
LOCATION/Qualifiers
source
1..1552
/organism="Solanum tuberosum"
/strain="Lemhi Russet"
/db_xref="taxon:4113"
82..1398
/gene="Potrall"
82..1398
/gene="Potrall"
/gene="Potrall"
/EC_number="2.2.1.2"
/codon_start=1
/product="transaldolase"
/protein_id="AA054016.1"
/db_xref="GI:2078350"
/translacion="MASICIKISNFCNENSTYGRSSSTPKFYVHLRTSPFNHLSAGS
KLSLVNPSIRASLCVKSQADNGSSVKKRTTHLHYETOGSPWMDNLCPLVTDILP
LIDSGVGVTSNPAIFOKAISTSNAYNDQFRELVOAKDIDSYWELVYNDLCLPVDACKL
FETLYDKTDGDDGVYSEVSPRLADDEGTVEAKMLKMKKERSNYIKIPATAPCIP
SIKEVVISIGISVNTLIESLSRYEAVIDAVLDGEASGLDLSRTVSASFVSRVDT
LVDKLLEKRTPEPNDLRKRAKNAQALADYDIOKRSRPMRAIVKGGKAKRLLMA
STSVKNPVPYPTLVDPIDPDTVSTPMDALDAFIDHGSVARTIDANLSEAGIYISA
LEKIDVWFVGSQLELGVDSEFKRSFDSLDLSLQERANTLKLIVNL"
BASE COUNT 439 a 315 c 343 g 455 t
ORIGIN

```

```

Query Match 14.5%; Score 38.2; DB 48; Length 1552;
Best Local Similarity 73.1%; Pred. No. 1.7;
Matches 49; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 181 ctcttaaaccttccttaacgatacgtcccaacagatggagatgcctgcaa 240
Db 239 CTATTAGGCGCATCTTGTGTGTCATATGCTCCCAAGCTGATGGAAGCTGATGA 298
QY 241 agagagac 247
Db 299 AGGAGAC 305
RESULT 3
LOCUS AC011452 147098 bp DNA HTG 18-JUL-2000
DEFINITION Homo sapiens chromosome 19 clone CTC-326K19, WORKING DRAFT
SEQUENCE, 3 ordered pieces.

```

ACCESSION AC011452
VERSION AC011452.5 GI:7711502
KEYWORDS HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 147098)
DOE Joint Genome Institute.
REFERENCE JOURNAL
AUTHORS Unpublished
TITLE Sequencing of Human Chromosome 19
DOE Joint Genome Institute.
REFERENCE JOURNAL
AUTHORS 2 (bases 1 to 147098)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 6, 2000 this sequence version replaced gi:7690094.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 310622, BC275645
Center Clone Name: CIT-HSPC_326K19

Summary Statistics
Consensus quality: 145829 bases at least Q40
Consensus quality: 146451 bases at least Q30
Consensus quality: 146700 bases at least Q20
Estimated insert size: 147530; agarose-fp estimation
Quality coverage: 6.66 in Q20 bases; agarose-fp estimation
Quality coverage: 6.66 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 22742: contig of 22742 bp in length
* 22743 22842: gap of unknown length
* 22843 80776: contig of 57934 bp in length
* 80777 80876: gap of unknown length
* 80877 147098: contig of 66222 bp in length.
FEATURES
Location/Qualifiers
1..147098
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone_lib="Caltech human BAC library C"
BASE COUNT 35229 a 37789 c 37738 g 36142 t 200 others
ORIGIN
Query Match 14.5%; Score 38.2; DB 39; Length 147098;
Best Local Similarity 61.6%; Pred. No. 1.4;
Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

AC010624/C
LOCUS 164313 bp DNA HTGS 21-JUN-2000
DEFINITION Homo sapiens chromosome 19 clone CTD-2126E3, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
ACCESSION AC010624.5 GI:8576043
VERSION AC010624.5
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 164313)
DOE Joint Genome Institute.
REFERENCE JOURNAL
AUTHORS Unpublished
TITLE Sequencing of Human Chromosome 19
DOE Joint Genome Institute.
REFERENCE JOURNAL
AUTHORS 2 (bases 1 to 164313)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jun 21, 2000 this sequence version replaced gi:7690069.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 676032, BC641056
Center Clone Name: CITB-HI_2126E3

Summary Statistics
Consensus quality: 157887 bases at least Q40
Consensus quality: 161345 bases at least Q30
Consensus quality: 162314 bases at least Q20
Estimated insert size: 212000; agarose-fp estimation
Quality coverage: 11.38 in Q20 bases; agarose-fp estimation
Quality coverage: 11.74 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1041: contig of 1041 bp in length
* 1042 1141: gap of unknown length
* 1142 5695: contig of 4554 bp in length
* 5696 5795: gap of unknown length
* 5796 14933: contig of 9138 bp in length
* 14934 15033: gap of unknown length
* 15034 25674: contig of 10641 bp in length
* 25675 25774: gap of unknown length
* 25775 44213: contig of 18439 bp in length
* 44214 44313: gap of unknown length
* 44314 61176: contig of 16863 bp in length
* 61177 61276: gap of unknown length
* 61277 164313: contig of 103037 bp in length.
FEATURES
Location/Qualifiers
1..164313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone_lib="Caltech human BAC library D"
BASE COUNT 43310 a 39759 c 39049 g 41593 t 602 others
ORIGIN
Query Match 14.5%; Score 38.2; DB 39; Length 164313;
Best Local Similarity 61.6%; Pred. No. 1.4;
Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

23775 29164: contig of 5390 bp in length

1. .163854

```
misc_feature 70903. 81217
              /note="assembly_fragment"
misc_feature 81318. 90922
              /note="assembly_fragment"
```

	a	c	g	t	others
BASE COUNT	50909	31168	30484	49055	2238
ORIGIN					

D_b 29909 CTCTTCCCTTCCTCTCTGTCCTCTCTCTCAAC 29871

SOURCE ORGANISM

REFERENCE AUTHORS

COMMENT

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L3000
Center clone name: 9_B.13

* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 854: contig of 854 bp in length
* 855 954: gap of 100 bp
* 955 1851: contig of 897 bp in length
* 1852 1951: gap of 100 bp
* 1952 2858: contig of 907 bp in length
* 2859 2958: gap of 100 bp
* 2959 3854: contig of 896 bp in length
* 3855 3954: gap of 100 bp
* 3955 4608: contig of 854 bp in length
* 4609 4908: gap of 100 bp
* 4909 5774: contig of 866 bp in length
* 5775 5874: gap of 100 bp
* 5875 6750: contig of 876 bp in length
* 6751 6850: gap of 100 bp
* 6851 7734: contig of 884 bp in length
* 7735 7834: gap of 100 bp
* 7835 8745: contig of 911 bp in length
* 8746 8845: gap of 100 bp
* 8846 9749: contig of 904 bp in length
* 9750 9849: gap of 100 bp
* 9850 10559: contig of 710 bp in length
* 10560 10659: gap of 100 bp
* 10660 11556: contig of 897 bp in length
* 11557 11656: gap of 100 bp
* 11657 12548: contig of 892 bp in length
* 12549 12648: gap of 100 bp
* 12649 13500: contig of 852 bp in length
* 13501 13600: gap of 100 bp
* 13601 14485: contig of 885 bp in length
* 14486 14585: gap of 100 bp
* 14586 15450: contig of 865 bp in length
* 15451 15550: gap of 100 bp
* 15551 16437: contig of 887 bp in length
* 16438 16537: gap of 100 bp
* 16538 17410: contig of 873 bp in length
* 17411 17510: gap of 100 bp
* 17511 18409: contig of 899 bp in length
* 18410 18509: gap of 100 bp
* 18510 19392: contig of 883 bp in length
* 19393 19492: gap of 100 bp
* 19493 20375: contig of 883 bp in length
* 20376 20475: gap of 100 bp
* 20476 21364: contig of 889 bp in length
* 21365 21464: gap of 100 bp
* 21465 22332: contig of 868 bp in length
* 22333 22432: gap of 100 bp
* 22433 23293: contig of 861 bp in length
* 23294 23393: gap of 100 bp
* 23394 24284: contig of 891 bp in length
* 24285 24384: gap of 100 bp
* 24385 25268: contig of 884 bp in length
* 25269 25368: gap of 100 bp
* 25369 26262: contig of 894 bp in length
* 26263 26362: gap of 100 bp
* 26363 27262: contig of 900 bp in length
* 27263 27362: gap of 100 bp
* 27363 28266: contig of 904 bp in length
* 28267 28366: gap of 100 bp

* 28367 29229: contig of 863 bp in length
* 29230 29329: gap of 100 bp
* 29330 30239: contig of 910 bp in length
* 30240 30339: gap of 100 bp
* 30340 31213: contig of 874 bp in length
* 31214 31313: gap of 100 bp
* 31314 32207: contig of 894 bp in length
* 32208 32307: gap of 100 bp
* 32308 33194: contig of 887 bp in length
* 33195 33294: gap of 100 bp
* 33295 34169: contig of 875 bp in length
* 34170 34269: gap of 100 bp
* 34270 35132: contig of 863 bp in length
* 35133 35232: gap of 100 bp
* 35233 36110: contig of 878 bp in length
* 36111 36210: gap of 100 bp
* 36211 37099: contig of 889 bp in length
* 37100 37199: gap of 100 bp
* 37200 38065: contig of 866 bp in length
* 38066 38165: gap of 100 bp
* 38166 39050: contig of 885 bp in length
* 39051 39150: gap of 100 bp
* 39151 40062: contig of 912 bp in length
* 40063 40162: gap of 100 bp
* 40163 41044: contig of 882 bp in length
* 41045 41144: gap of 100 bp
* 41145 42060: contig of 916 bp in length
* 42061 42160: gap of 100 bp
* 42161 43051: contig of 891 bp in length
* 43052 43151: gap of 100 bp
* 43152 44041: contig of 890 bp in length
* 44042 44141: gap of 100 bp
* 44142 44996: contig of 855 bp in length
* 44997 45096: gap of 100 bp
* 45097 45958: contig of 863 bp in length
* 45960 46059: gap of 100 bp
* 46060 46945: contig of 886 bp in length
* 46946 47045: gap of 100 bp
* 47046 47946: contig of 901 bp in length
* 47947 48046: gap of 100 bp
* 48047 48913: contig of 873 bp in length
* 48920 49019: gap of 100 bp
* 49020 49920: contig of 901 bp in length
* 49921 50020: gap of 100 bp
* 50021 50898: contig of 878 bp in length
* 50899 50998: gap of 100 bp
* 50999 51891: contig of 893 bp in length
* 51892 51991: gap of 100 bp
* 51992 52893: contig of 902 bp in length
* 52894 52993: gap of 100 bp
* 52994 53889: contig of 866 bp in length
* 53890 53989: gap of 100 bp
* 53990 54867: contig of 878 bp in length
* 54868 54967: gap of 100 bp
* 54968 55811: contig of 844 bp in length
* 55812 55911: gap of 100 bp
* 55912 56786: contig of 875 bp in length
* 56787 56886: gap of 100 bp
* 56887 57787: contig of 901 bp in length
* 57788 57887: gap of 100 bp
* 57888 58783: contig of 886 bp in length
* 58784 58883: gap of 100 bp
* 58884 59770: contig of 887 bp in length
* 59771 59870: gap of 100 bp
* 59871 60782: contig of 912 bp in length
* 60783 60882: gap of 100 bp
* 60883 61771: contig of 889 bp in length
* 61772 61871: gap of 100 bp
* 61872 62779: contig of 908 bp in length
* 62780 62879: gap of 100 bp
* 62880 63759: contig of 880 bp in length
* 63760 63859: gap of 100 bp
* 63860 64764: contig of 905 bp in length

[illegible]

note: this is a working draft sequence. It currently consists of 15 collaps. The time order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the collaps are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number willl be preserved. 5007

```
misc_feature 1. .6987 /note="assembly_fragment:00213000
clone_end:SP6
vector_side:left"
7088. .25926 /note="assembly_fragment:00622
fragment_chain:1"
26027. .27279 /note="assembly_fragment:00813000
fragment_chain:1"
27380. .44222 /note="assembly_fragment:00646666
fragment_chain:1"
44333. .69220 /note="assembly_fragment:00797979
fragment_chain:1"
69321. .81427 /note="assembly_fragment:015311
fragment_chain:1"
81528. .105379 /note="assembly_fragment:013000
fragment_chain:1"
103480. .129641 /note="assembly_fragment:001266
fragment_chain:1"
129742. .139328 /note="assembly_fragment:01328
```

[illegible]

	RESULT	12
LOCUS	AC026928	
DEFINITION	AC026928 187999 bp DNA HTG	22-JUL-2000
VERSION	Homo sapiens chromosome 9 clone RP11-18A3 map 9,	WORKING DRAFT
SOURCE	SEQUENCE, 37 unordered pieces.	
ORGANISM	AC026928.2 GI:9369514 HTG; HTGS_PHASE1; HTGS_DRAFT. human.	
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebratele; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 187999) Birren,B., Linton,L., Nusbbaum,C. and Lander,E.	
TITLE	Unpublished	
JOURNAL	2 (bases 1 to 187999)	
AUTORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F., Boudslavsky,I., Boukhalter,B., Brown,A., Burkett,G., Campomali,A., Castle.A., Choepel,Y., Colangelo,M., Collins.S., Collamore,A., Cooke,P., Dekrellano,K., Dewar.K., Diaz.J.S., Dodge,S., Domino.M., Doyle,M., Ferreira.P., FitzHugh.W., Gage.D., Galagan,U., Gardyna,S., Ginde,S., Goylette,M., Graham.L., Grand-Pierre.N., Grant.G., Hagos.B., Heatford.A., Horton.L., Howland,D.C., Iliev.I., Johnson.R., Jones.C., Kamp.L., Karlas.A., Klein,J., LaRocque,K., lamazaras,R., Landers,T., Lehoczy,U., Leyne.R., Lien.C., Liu.G., Locke,K., MacDonald,P., Margulis,N., McCarthy.M., McEwan,P., McGurk,A., McKernan,K., MCPheeters,R., Melidirm.U., Meunus.L., Mihova.T., Miranda.C., Mlenaga.V., Morrow,J., Murphy.T., Naylor.J., Norman.C.H., O'Connor,T., O'Donnell.P., O'Neill,D., Oliver.T.M., Oliver.J., Peterson,K., Pierre.N., Pisanti.C., Pollara.V., Raymond.C., Riley.R., Rogov,P., Rofman.D., Roy.A., Santos.R., schauer.S., Severy,P., Spencer.B.,	

TITLE
JOURNAL
COMMENT

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teifaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 22, 2000 this sequence version replaced g1:7328793.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 18_A3
Center clone name: 18_A3

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167725 bases at least Q40
Consensus quality: 178025 bases at least Q30
Consensus quality: 182047 bases at least Q20
Insert size: 194000; agarose- ϕ p
Insert size: 184399; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose- ϕ p
Quality coverage: 4.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1227: contig of 1227 bp in length
* 1228 1327: gap of 100 bp
* 1328 2416: contig of 1089 bp in length
* 2417 2516: gap of 100 bp
* 2517 4057: contig of 1541 bp in length
* 4058 4157: gap of 100 bp
* 4158 5171: contig of 1014 bp in length
* 5172 5271: gap of 100 bp
* 5272 6380: contig of 1109 bp in length
* 6381 6480: gap of 100 bp
* 6481 7659: contig of 1179 bp in length
* 7660 7759: gap of 100 bp
* 7760 9214: contig of 1455 bp in length
* 9215 9314: gap of 100 bp
* 9315 10932: contig of 1618 bp in length
* 10933 11032: gap of 100 bp
* 11033 12035: contig of 1003 bp in length
* 12036 12135: gap of 100 bp
* 12136 13215: contig of 1080 bp in length
* 13216 13315: gap of 100 bp
* 13316 14645: contig of 1330 bp in length
* 14646 14745: gap of 100 bp
* 14746 15925: contig of 1180 bp in length
* 15926 16025: gap of 100 bp
* 16026 17158: contig of 1133 bp in length
* 17159 17258: gap of 100 bp
* 17259 19445: contig of 2187 bp in length
* 19446 19545: gap of 100 bp
* 19546 21242: contig of 1697 bp in length
* 21243 21342: gap of 100 bp
* 21343 24035: contig of 2693 bp in length
* 24036 24135: gap of 100 bp
* 24136 25142: contig of 1007 bp in length
* 25143 25242: gap of 100 bp

FEATURES

source

25243 27117: contig of 1875 bp in length
* 27118 27217: gap of 100 bp
* 27218 28912: contig of 1655 bp in length
* 28913 29012: gap of 100 bp
* 29013 31508: contig of 2456 bp in length
* 31509 31608: gap of 100 bp
* 31609 32697: contig of 1089 bp in length
* 32698 32798: gap of 100 bp
* 32798 37275: contig of 4478 bp in length
* 37276 37375: gap of 100 bp
* 37376 40980: contig of 3605 bp in length
* 40981 41080: gap of 100 bp
* 41081 45215: contig of 4135 bp in length
* 45216 45315: gap of 100 bp
* 45316 48710: contig of 3395 bp in length
* 48711 48810: gap of 100 bp
* 48811 52850: contig of 4040 bp in length
* 52851 52950: gap of 100 bp
* 52951 57072: contig of 4122 bp in length
* 57073 57172: gap of 100 bp
* 57173 62417: contig of 5245 bp in length
* 62418 62517: gap of 100 bp
* 62518 71007: contig of 8490 bp in length
* 71008 71107: gap of 100 bp
* 71108 79805: contig of 8698 bp in length
* 79806 79905: gap of 100 bp
* 79906 81160: contig of 1255 bp in length
* 81161 81260: gap of 100 bp
* 81261 92353: contig of 11093 bp in length
* 92354 92453: gap of 100 bp
* 92454 101125: contig of 8672 bp in length
* 101126 101225: gap of 100 bp
* 101226 117073: contig of 15848 bp in length
* 117074 117173: gap of 100 bp
* 117174 136020: contig of 18847 bp in length
* 136021 136120: gap of 100 bp
* 136121 161828: contig of 25708 bp in length
* 161829 161928: gap of 100 bp
* 161929 187999: contig of 26071 bp in length.
Location/Qualifiers
1. 187999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="g"
/clone="RP11-18A3"
/clone_id="RPC1-11 Human Male BAC"
1. 1227
/note="assembly_fragment"
misc_feature
1328. 2416
/note="assembly_fragment"
2517. 4057
/note="assembly_fragment"
4158. 5171
/note="assembly_fragment"
5272. 6380
/note="assembly_fragment"
6481. 7659
/note="assembly_fragment"
7760. 9214
/note="assembly_fragment"
9315. 10932
/note="assembly_fragment"
11033. 12035
/note="assembly_fragment"
12136. 13215
/note="assembly_fragment"
13316. 14645
/note="assembly_fragment"
14746. 15925
/note="assembly_fragment"
16026. 17158
/note="assembly_fragment"

Query Match	13.9%	Score 36.8;	DB 54;	Length 187999;
Best Local Similarity	53.5%;	Pred. No. 3.3;		
Matches 77; Conservative	0;	Mismatches 67;	Indels 0;	Gaps 0;

Db 119379 CCCCTCAATACCTAGCAGTCATCTCTCTCTCCATGCCATTTCCTCTGTCTATT 119438
QY 125 cgccttaacacaacatacctatccgttcctcccgatccaattactgctcctc 164

RESULT 13	AC022699/c	LOCUS	DEFINITION
	AC022699	194963 bp	DNA
		Mus musculus clone RP3-10718,	HTG
		WORKING DRAFT SEQUENCE, 38 unretarded	07-JUL-2000

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
1 (bases 1 to 194963)	Birren, B., Linton, L., Nusbaum, C. and Lander, E.	Mus musculus, clone RP23-10718	Unpublished	2 (bases 1 to 194963)	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abramson, H., Allen, N.

Anderson, S., Baldwin, J., Barna, N., Beckley, E., Buchanan, H., Allen, N., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A., Chopeel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deatelliano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gadyana, S., Grant, G., Hagos, B., Heatford, A., Horton, I., Howland, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J., Macdonald, P., Margolis, N., Levine, R., Lieu, C., Liu, G., Locke, K., McPheters, R., Melgrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K., Pierre, N., Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rotman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talence, J., Teedavay, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,

45520 45619: gap of 100 bp

* 45620 50053: contig of 4434 bp in length

```

* 50054 50153: gap of 100 bp
* 50154 55501: contig of 5348 bp in length
* 55502 55601: gap of 100 bp
* 55602 60490: contig of 4889 bp in length
* 60491 60590: gap of 100 bp
* 60591 66444: contig of 5854 bp in length
* 66445 66544: gap of 100 bp
* 66545 71111: contig of 4567 bp in length
* 71112 71211: gap of 100 bp
* 71212 77867: contig of 6656 bp in length
* 77868 77967: gap of 100 bp
* 77968 82747: contig of 4780 bp in length
* 82748 82847: gap of 100 bp
* 82848 88691: contig of 5844 bp in length
* 88692 88791: gap of 100 bp
* 88792 95730: contig of 6939 bp in length
* 95731 95830: gap of 100 bp
* 95831 103037: contig of 7207 bp in length
* 103038 103137: gap of 100 bp
* 103138 111460: contig of 8323 bp in length
* 111461 111560: gap of 100 bp
* 111561 119148: contig of 7588 bp in length
* 119149 119248: gap of 100 bp
* 119249 128097: contig of 8849 bp in length
* 128098 128197: gap of 100 bp
* 128198 137747: contig of 9550 bp in length
* 137748 137847: gap of 100 bp
* 137848 149328: contig of 11481 bp in length
* 149329 149428: gap of 100 bp
* 149429 160664: contig of 11236 bp in length
* 160665 160764: gap of 100 bp
* 160765 170402: contig of 9638 bp in length
* 170403 170502: gap of 100 bp
* 170503 182201: contig of 11699 bp in length
* 182202 182301: gap of 100 bp
* 182302 194963: contig of 12662 bp in length.

```

FEATURES

source

```

misc_feature 1..1103
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /clone_lib="RPCI-23 Female Mouse BAC"
misc_feature 1204..2643
             /note="assembly_fragment"
misc_feature 2744..3758
             /note="assembly_fragment"
misc_feature 3859..4988
             /note="assembly_fragment"
misc_feature 5089..6259
             /note="assembly_fragment"
misc_feature 6360..8126
             /note="assembly_fragment"
misc_feature 8227..9555
             /note="assembly_fragment"
misc_feature 9656..11586
             /note="assembly_fragment"
misc_feature 11687..13441
             /note="assembly_fragment"
misc_feature 13542..14986
             /note="assembly_fragment"
misc_feature 15087..18988
             /note="assembly_fragment"
misc_feature 19089..21553
             /note="assembly_fragment"
misc_feature 21654..24665
             /note="assembly_fragment"
misc_feature 24766..29058
             /note="assembly_fragment"
misc_feature 29159..31541
             /note="assembly_fragment"
misc_feature 31642..35036
             /note="assembly_fragment"

```

```

misc_feature 35137..38278
             /note="assembly_fragment"
misc_feature 38379..41421
             /note="assembly_fragment"
misc_feature 41522..45519
             /note="assembly_fragment"
misc_feature 45620..50053
             /note="assembly_fragment"
misc_feature 50154..55501
             /note="assembly_fragment"
misc_feature 55602..60490
             /note="assembly_fragment"
misc_feature 60591..66444
             /note="assembly_fragment"
misc_feature 66545..71111
             /note="assembly_fragment"
misc_feature 71212..77867
             /note="assembly_fragment"
misc_feature 77968..82747
             /note="assembly_fragment"
misc_feature 82848..88691
             /note="assembly_fragment"
misc_feature 88792..95730
             /note="assembly_fragment"
misc_feature 95731..95830
             /note="assembly_fragment"
misc_feature 95831..103037
             /note="assembly_fragment"
misc_feature 103038..103137
             /note="assembly_fragment"
misc_feature 103138..111460
             /note="assembly_fragment"
misc_feature 111461..111560
             /note="assembly_fragment"
misc_feature 111561..119148
             /note="assembly_fragment"
misc_feature 119149..119248
             /note="assembly_fragment"
misc_feature 119249..128097
             /note="assembly_fragment"
misc_feature 128098..128197
             /note="assembly_fragment"
misc_feature 128198..137747
             /note="assembly_fragment"
misc_feature 137748..137847
             /note="assembly_fragment"
misc_feature 137848..149328
             /note="assembly_fragment"
misc_feature 149329..149428
             /note="assembly_fragment"
misc_feature 149429..160664
             /note="assembly_fragment"
misc_feature 160665..160764
             /note="assembly_fragment"
misc_feature 160765..170402
             /note="assembly_fragment"
misc_feature 170403..170502
             /note="assembly_fragment"
misc_feature 170503..182201
             /note="assembly_fragment"
misc_feature 182202..182301
             /note="assembly_fragment"
misc_feature 182302..194963
             /note="assembly_fragment"

```

Query Match 13.9% Score 36.8; DB 51; Length 194963;
 Best Local Similarity 56.7%; Pred. No. 3.3;
 Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```

Oy 28 ccgttcccaagctctcaacgcaatccacttgccttcattacgaagcgttcgctccc 87
    || || || || || || || || || || || || || || || || || || || ||
Db 84605 CAGCTTCCCAAGTGTCAAGAACCAAGCAGTGTCCCTTGGTGTGCTGTGCTTC 84546
    || || || || || || || || || || || || || || || || || || || ||
Oy 88 gagattccgcttctcctcctcctcaaccttccatcgctttatcacaaactaac 147
    || || || || || || || || || || || || || || || || || || || ||
Db 84545 CCTCTCACCAAGTCTCACTTCCGCAAGTCTCCCGTCCGTCTGACGACGCGCCAAAC 84486
    || || || || || || || || || || || || || || || || || || || ||

```

RESULT 14

HSL190H5

LOCUS HSL190H5 24721 bp DNA PRI 23-NOV-1999
 DEFINITION Human DNA sequence from cosmid L190H5, Huntington's Disease Region,
 chromosome 4p16.3.

ACCESSION

VERSION Z68163.1 GI:1122279
 KEYWORDS 4p16.3; CPG island; repeat polymorphism.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

Mungall, A.
 Direct Submission
 Submitted (05-DEC-1995) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1RQ, England. E-mail enquires: humpub@sanger.ac.uk

JOURNAL

COMMENT

On Dec 14, 1995 this sequence version replaced gi:1107691.
 IMPORTANT:
 This sequence is not the entire insert of clone L190H5. It may be
 shorter because we only sequence overlapping sections once, or
 longer because we arrange for a small overlap between neighbouring
 submissions.

The three left end of clone L190H5 is at 1 in this sequence. The
 true left end of clone L129H7 is at 24622 in this sequence. L190H5
 is from cosmid library LAD4NCO1 constructed at the Human Genome
 Center, Los Alamos National Laboratory, NM 87545 under the auspices
 of the U.S. Department of Energy. The library was constructed using
 flow-sorted human chromosome 4 from a Hamster-Human hybrid cell
 line (UV20H121-27) containing human chromosomes 4, 8 and 21.

VECTOR: scos1

L190H5 is contained in a clone contig spanning
 2mb which is described in Baxendale et al, Nature Genetics 4 (1993)
) 181-186. See also Myers et al, Cytogenet Cell Genet. 66 (1994)
) 218-230.

FEATURES

Location/Qualifiers


```
* 7143 15239: contig of 8097 bp in length
* 15240 15339: gap of 100 bp
* 15340 40312: contig of 24973 bp in length
* 40313 40412: gap of 100 bp
* 40413 41496: contig of 1084 bp in length
* 41497 41596: gap of 100 bp
* 41597 45493: contig of 3897 bp in length
* 45494 45593: gap of 100 bp
* 45594 81525: contig of 35932 bp in length
* 81526 81625: gap of 100 bp
* 81626 105239: contig of 23614 bp in length
* 105240 105339: gap of 100 bp
* 105340 132375: contig of 27036 bp in length
* 132376 132475: gap of 100 bp
* 132476 142831: contig of 10356 bp in length
* 142832 142931: gap of 100 bp
* 142932 144114: contig of 1183 bp in length
* 144115 144214: gap of 100 bp
* 144215 156872: contig of 12658 bp in length
* 156873 156972: gap of 100 bp
* 156973 181517: contig of 24545 bp in length.
```

FEATURES

SOURCE

```
1. 181517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-574M5"
/clone_1lb="RPC1-11.2"
1. 7042
/note="assembly-fragment:00172"
clone_end:"7"
vector_side:"left"
7143. 15239
/note="assembly-fragment:00673"
fragment_chain:1"
15340. 40312
/note="assembly-fragment:00607"
fragment_chain:1"
40413. 41496
/note="assembly-fragment:00878"
fragment_chain:1"
41597. 45493
/note="assembly-fragment:01120"
fragment_chain:2"
45594. 81525
/note="assembly-fragment:00551"
fragment_chain:2"
81626. 105239
/note="assembly-fragment:00676"
fragment_chain:2"
105340. 132375
/note="assembly-fragment:00357"
fragment_chain:3"
132476. 142831
/note="assembly-fragment:00439"
fragment_chain:3"
142932. 144114
/note="assembly-fragment:00325"
144215. 156872
/note="assembly-fragment:00621"
156973. 181517
/note="assembly-fragment:00649"
BASE COUNT 54727 a 36050 c 35860 g 53777 t 1103 others
ORIGIN
```

Query Match

13.9%; Score 36.6; DB 68; Length 181517;

Best Local Similarity 53.1%; Pred. No. 3.8;

Matches --78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 52 atccactgtcttcattcagaagcgcttcgccccgagatctcgtcttcacttca 111

.Db 129772 ATCTCTGACCTCGTATCCGCCGCTTGTGCTCCAGAGTGTGGATTCAGAGTGTG 129713

QY 112 aactcttcacatgcctttaacacaaactacatcgccttcctccatccattc 171

.Db 129712 AGCCACCGTGCCTGGCTGTGACGTCACTCTTTAACCACTCACTTCCACCAACCAAT 129653

QY 172 actgccttccttaaaactccttac 198

.Db 129652 ACCGCTCCACCAACCAACCAACCAAC 129626

Search completed: November 11, 2000, 18:38:17
Job time: 1583 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2000, 18:15:50 ; Search time 74.01 Seconds

(without alignments)
1340.021 Million cell updates/sec

Title: US-09-300-482-569

Perfect score: 264
Sequence: 1 ctgactgcagcatccca.....gacatgctcatgacttta 264

Scoring table: IDENTITY_NNC
Gapop 10.0 , Gapext 1.0

Searched: 480022 segs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID6/gcgdata/geneseq/NA1980.DAT.*
2: /SID6/gcgdata/geneseq/NA1981.DAT.*
3: /SID6/gcgdata/geneseq/NA1982.DAT.*
4: /SID6/gcgdata/geneseq/NA1983.DAT.*
5: /SID6/gcgdata/geneseq/NA1984.DAT.*
6: /SID6/gcgdata/geneseq/NA1985.DAT.*
7: /SID6/gcgdata/geneseq/NA1986.DAT.*
8: /SID6/gcgdata/geneseq/NA1987.DAT.*
9: /SID6/gcgdata/geneseq/NA1988.DAT.*
10: /SID6/gcgdata/geneseq/NA1989.DAT.*
11: /SID6/gcgdata/geneseq/NA1990.DAT.*
12: /SID6/gcgdata/geneseq/NA1991.DAT.*
13: /SID6/gcgdata/geneseq/NA1992.DAT.*
14: /SID6/gcgdata/geneseq/NA1993.DAT.*
15: /SID6/gcgdata/geneseq/NA1994.DAT.*
16: /SID6/gcgdata/geneseq/NA1995.DAT.*
17: /SID6/gcgdata/geneseq/NA1996.DAT.*
18: /SID6/gcgdata/geneseq/NA1997.DAT.*
19: /SID6/gcgdata/geneseq/NA1998.DAT.*
20: /SID6/gcgdata/geneseq/NA1999.DAT.*
21: /SID6/gcgdata/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	32.4	12.3	857	20	X57401 Rat U3 gene trap d
2	31.4	11.9	1612	19	V44595 Human respiration
3	31.4	11.9	4002	21	Z60603 Nucleotide sequenc
4	31.2	11.8	53526	19	T94101 Human PKD1 gene.
5	31.2	11.8	53577	17	T18551 Human polycystic k
6	31.2	11.8	53577	19	T94108 Human PKD1 locus
7	30.8	11.7	36901	20	Z23897 Murine LOBO genom
8	30.8	11.7	38886	20	Z23897 Murine LOBO genom
c 9	29.8	11.3	1798	18	T51756 DAX-1 gene promot
c 10	29.6	11.2	1036	18	T19421 M. tuberculosis an
11	29.6	11.2	1036	20	Z19209 M. tuberculosis re
c 12	29.6	11.2	1664976	19	V21209 Methanococcus jann

13	29.4	11.1	3935	21	Z89458
c 14	29.2	11.1	12793	19	V44251
c 15	29	11.0	1005	20	Z16507
c 16	29	11.0	5752	20	Z30052
c 17	28.8	10.9	525	18	T67902
c 18	28.8	10.9	1444	21	Z93353
c 19	28.8	10.9	2575	16	O89558
c 20	28.8	10.9	2575	16	O90331
c 21	28.6	10.8	433	20	X52010
c 22	28.6	10.8	549	21	A15937
c 23	28.6	10.8	696	20	X97697
c 24	28.6	10.8	966	21	Z36244
c 25	28.6	10.8	997	21	A15947
c 26	28.6	10.8	2388	20	V84554
c 27	28.4	10.7	910715	20	X20248
c 28	28.2	10.7	1239	21	Z44078
c 29	28.2	10.7	1361	21	A02089
c 30	28.2	10.7	1451	21	Z56720
c 31	28	10.6	1593	21	A02504
c 32	28	10.6	2253	10	N92071
c 33	27.8	10.5	614	21	Z80128
c 34	27.8	10.5	1496	21	Z52486
c 35	27.6	10.5	758	19	V44215
c 36	27.6	10.5	1235	20	Z16147
c 37	27.6	10.5	1442	16	O90211
c 38	27.6	10.5	2035	18	V30292
c 39	27.6	10.5	3131	18	T91708
c 40	27.6	10.5	3562	19	V60839
c 41	27.6	10.5	3562	21	Z61197
c 42	27.6	10.5	3647	19	V60842
c 43	27.6	10.5	3647	21	Z61200
c 44	27.6	10.5	18443	20	V83942
c 45	27.4	10.4	420	21	Z80329

ALIGNMENTS

RESULT 1	
X57401/c	
ID X57401 standard; DNA; 857 BP.	
XX	
AC X57401:	
XX	
DT 24-JUL-1999 (first entry)	
XX	
DE Rat U3 gene trap derived nucleic acid 14A13E-RE.	
XX	
KW Gene trap; rat; cellular gene; viral infection; cell survival; cancer;	
KW tumour progression; suppression; identification; viral growth;	
KW tumour suppressor; prevention; screening; therapeutic agent; ss.	
XX	
OS Rattus norvegicus.	
XX	
PN W09919481-A2.	
XX	
PD 22-APR-1999.	
XX	
PF 08-OCT-1998: 98WO-US21276.	
XX	
PR 10-OCT-1997: 97US-0062021.	
XX	
PA (UYVA-) UNIV VANDERBILT.	
XX	
PI Dubois RN, Organ EL, Rubin DH;	
XX	
DR WPI; 1999-326546/27.	
XX	
PT Nucleic acid encoding tumor suppressors and products required for	
PT viral infection	
XX	
PS Claim 1; Page 58-59; 94pp; English.	

CC This invention describes novel rat-derived nucleic acid fragments from
CC cellular genes that are necessary for viral infection but not for cell
CC survival, or that suppress tumour progression. The products of the
CC invention (A57371-1, A57497) can be used in methods of identifying cellular
CC genes necessary for viral growth and cellular genes that function as
CC tumour suppressors and for predicting or preventing such infections or
CC cancer. They may also be used in screening for potential therapeutic
CC agents. These sequences can be targeted without significant side effects
CC (contrast targeting genes essential for viral growth).

SQ Sequence 857 BP; 189 A; 151 C; 234 G; 220 T; 63 other,

	Query Match	Similarity	Score	DB	Length
Best	Local	47.5%	Pred. No.	0.28	
Matches	66	Conservative	0	Mismatches	73
				Indels	0
				Gaps	0

	Query	Match	Similarity	Score	DB	Length
QY	62	ttcctcatcagaagcgcgtcgcgccccgagatctcgcgtctcctaacttaaacctcttc	121			
DB	248	TTGCNAGATNAGAACTNCANCTTCTTCANCGTNGTGTCCCAACCCNTTCAACNCGCTNCG	189			
QY	122	catgcgtttatcacaacaactaactatcgcgcttcctccgcgatcattcgcgtcc	181			
DB	188	TTTTCCGTTTGGCNCACACCAACNTTTCCTTCCCTTANTCNCGCCGACCCGTTTCNTGNC	129			
QY	182	tcttaaacctcctciacg	200			
DB	128	NCTGGAAATTTTTCNCG	110			

RESULT	2
V44595	
ID	V44595 standard; cDNA; 1612 BP.

DT 24-NOV-1998 (first entry)

DE Human respiration uncoupling protein-2 (UCP2) cDNA.

KM uncoupling protein-2, ucp2 gene; human: respiration;
KM thermogenesis; obesity; hyperinsulinaemia; glucose intolerance;
KM diabetes mellitus; wasting; cachexia; anorexia;
KM diarrhoea; syndrome x; hypothermia; gene therapy; diagnosis; ss
KM inflammation; fever; hyperthermia; gene therapy; diagnosis; ss

05 Homo sapiens.

Key	Location/Qualifiers
EH	1074

ET

ET

FT variation

ET

FT 13

FT variation

ET

ET 2

FT variation

FT FT

ET ET

FT variation
FT

10

FT	/note=	"C to T base change"
FT	1224	
FT	variation	
FT	/*tyg=	j
FT	/note=	"C to A base change (L294M)"
FT	1262	
FT	variation	
FT	/*tyg=	k
FT	/note=	"G to A base change"

PN W09831396-A1

PD 23-JUL-1998

PF 22-APR-1997; 97WO-US06864.

PR 15-JAN-1997; 97US-0034960.

PA (NARE-) CENT NAT RECH SCI CENT RECH SUR ENDOCRINOLOGIA
PA (RECC) UNIV CALTEORNTA

PA (UYDU-) UNIV DUKE.
 YX

PI Bouillaud F, Collins SA, Riequier D, Serain M,
 RI Surwit RS, Warden CH:

XX
WPT. 1998-41382

P-PSDB; W69166.
DR
XX

Method for treating disease associated with increased UCP-2 activity, by administering agent which enhances or inhibits UCP-2 activity,

PT effectively t
cachexia etc
PT

XX Disclosure; Fig 1e; 98pp; English
PS

This is a consensus sequence for the entire human expressed UCP2 (mRNA). It includes a coding region for a novel protein (see W69166), designated uncoupling protein-2. Clones hUCP2-5' (ATCC 97850) and hUCP2-3' (ATCC 97849), from which the consensus is derived, were isolated from a human lung Marathon library (Clontech) by RACE using primers hUCP2.CDS3 and hUCP2.CDS3 (V44623-24). A number of polymorphisms have been identified in the sequence. The UCP2 gene maps to a chromosomal region (11q13) linked to obesity and hyperinsulinaemia. A claimed method for treating a disease or disorder associated with diminished (or elevated) UCP2 expression or activity comprises administering an agent that enhances (or inhibits) UCP2 expression or activity sufficient to effect the treatment. Also claimed are: (1) a method for screening a compound for its ability to modulate UCP2 expression or activity; (2) isolated mammalian UCP2; (3) an isolated nucleic acid sequence encoding mammalian UCP2; (4) a recombinant molecule comprising a vector and the nucleic acid of (3); (5) a host cell that can be used to produce UCP2; (6) an antibody specific for UCP2; (7) an antisense construct complementary to UCP2; and (8) a host cell comprising the construct of (7). An agent which enhances UCP2 expression can be used to treat obesity, diabetes, syndrome X, hypothermia, hyperinsulinaemia, or glucose intolerance. An inhibitor of UCP2 is used to treat wasting, anorexia, inflammation, cachexia, fever or hyperthermia. The constructs of (4) and (7) can both be used in the method (all claimed). The invention also relates to diagnostic methodologies.

Sequence 1612 BP; 307 A; 517 C; 394 G; 394 T; 0 other;

Query Match	11.9%	Score 31.4	DB 19	Length 1612
Best Local Similarity	51.08	Pred. No. 0.75		
Matches 74	Conservative	0	Mismatches 71	Indels 0
				Gaps 0

[illegible]

Db 1373 ctccctcttcccccactctctctccggtcccttaaccacactctcttc 1432
OY 159 ctccgatccattcactgctccctc 183
Db 1433 aacattcactcactcacttgc 1457

RESULT 3

260603
ID 260603 standard; DNA: 4002 BP.

XX 260603;

XX 16-MAY-2000 (first entry)

XX Nucleotide sequence encoding a human Sos protein.

XX Ras; Son of sevenless; Sos; crystal; Ras-Sos complex; cancer;

XX protein coordinate data; ss.

XX Homo sapiens.

XX MO200005258-A1.

XX 03-FEB-2000.

XX 20-JUL-1999; 99WO-US16348.

XX 21-JUL-1998; 98US-0119794.

XX (UYRQ) UNIV ROCKEFELLER.

XX Borlack-Sjodin A, Margarit SM, Bar-Sagi D, Cole P, Kuriyan J;

XX WPI: 2000-182647/16.

XX Novel crystals comprising a Ras-Son of sevenless complex, useful for
XX screening drugs useful in cancer treatment -

XX Disclosure: Page 197-198; 224pp; English.

XX The specification describes a crystal complex comprising at least a
XX Sos contacting region of a Ras protein and at least a Son of sevenless
XX (Sos) protein catalytic region fragment, that effectively diffracts
XX x-rays. Ras and Sos form a tight complex. Sos does not impede the
XX binding sites for the nucleotide base and the ribose of GTP or GDP
XX and thus the Ras-Sos complex maintains a structure that permits
XX nucleotide release and rebinding. The crystals are used for the
XX determination of the atomic coordinates of the complex to a resolution
XX of more than 5.0 Angstrom. The crystals, or a dataset comprising the
XX three-dimensional coordinates obtained from the crystals, is useful
XX for identifying an agent that stabilizes the Ras-Sos complex. The
XX crystals are also useful for identifying agents that inhibit the
XX formation of Ras-Sos complex. Ras and Sos fragments are useful for
XX growing a crystal of a protein-ligand complex. Agents that stabilize
XX or inhibit the formation of Ras-Sos complex are useful in the
XX treatment of cancer. The present sequence encodes a Sos protein.

XX Sequence 4002 BP; 1341 A; 829 C; 783 G; 1049 T; 0 other;

XX Query Match 11.9%; Score 31.4; DB 21; Length 4002;

XX Best Local Similarity 54.9%; Pred. No. 1.1; Mismatches 51; Indels 0; Gaps 0;

XX Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 83 ccccgagattcgtcttcacttaaacctcttcacatgctttaaacaac 142

Db 3665 caccgagatttctcaagcaccatacctccaaccccttgggcaaaaaa 3724

OY 143 taactatcgcttccctccgatccatccactcctcttaaacctct 195

Db 3725 gtgacacagcaagctctctcccaaacagccttcccttaccacact 3777

RESULT 4

ID T94101 standard; DNA: 53526 BP.

XX T94101;

XX 01-JUN-1998 (first entry)

XX Human PKD1 gene.

XX Human; polycystic kidney disease 1; PKD1; treatment;

XX autosomal dominant polycystic kidney disease; APKD; ss.

XX Homo sapiens.

XX WO9744457-A1.

XX 27-NOV-1997.

XX 22-MAY-1997; 97WO-US08799.

XX 03-JUN-1996; 96US-0658136.

XX 24-MAY-1996; 96US-0655360.

XX (GENZ) GENZYME CORP.

XX Burn T, Connors T, Dackowski W, Germino G, Klinger K;

XX Qian F;

XX WPI: 1998-018511/02.

XX Human polycystic kidney disease gene, PKD1 - useful to treat and
XX diagnose human autosomal or adult onset polycystic kidney disease

XX Claim 2: Pages 90-118; 257pp; English.

XX The present sequence is the human polycystic kidney disease 1
XX (PKD1) gene. The PKD1 gene or polypeptide may be used to treat
XX autosomal dominant polycystic kidney disease (APKD), and identify
XX carriers of mutant PKD1 genes, i.e. subjects susceptible to APKD.
XX Antibodies (Ab) that distinguish between normal and mutant PKD1
XX sequences can also be used in diagnostic tests. Anti-PKD1 Ab may
XX also be used to perform subcellular and histochemical localisation
XX studies, and to block the function of PKD1. Ab are also useful in
XX rational drug design studies to identify and test inhibitors of
XX PKD1. Sense and antisense sequences derived from the PKD1 gene may
XX used for detection and therapy.

XX Sequence 53526 BP; 8486 A; 17665 C; 15768 G; 11607 T; 0 other;

XX Query Match 11.8%; Score 31.2; DB 19; Length 53526;

XX Best Local Similarity 47.4%; Pred. No. 3.3; Mismatches 103; Indels 0; Gaps 0;

XX Matches 93; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

OY 18 ccaatgcttcggttcccaagcttccaagccaatccactgtctcttaccagaagc 77

Db 35274 cccatgccccctccctccctccctccctccctccctccctccctccctcc 35333

OY 78 ctggccccgagattcgtcttcacttcaaccttccatcccttccatcgaagac 137

Db 35334 atccccctctctctccctctctctccctctctctctctctctctctctct 35393

OY 138 aaaaactaactacgcgttcctccctccgatccatccactcgtctctttaaactccta 197

Db 35394 ctctccctccctccatccatccctccctccatccatccctccctccatccatcc 35453

OY 198 cggatcaaatgtctcc 213

Db 35454 cctcccatctctc 35469

1T18551	5	Result
ID	T18551	standard; DNA; 53577 BP.
AC	T18551;	
XX	06-MAY-1997	(first entry)
DE	Human polycystic kidney disease normal PKD1 gene.	
XX	Adult onset polycystic kidney disease; APKD; autosomal dominant;	
KW	mutant; transversion; transition; deletion; insertion; ds.	
XX	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	misc_feature	4379..5272
FT	/*tag= a	/note= "specifically claimed region of intronless
FT	/*tag= b	CDNA identified by exon trapping"
FT	old_sequence	replace(50652..50653, cg)
FT	old_sequence	/note= "changes Val codon to Leu codon"
FT	old_sequence	replace(50796..50797, cg)
FT	/*tag= c	/note= "replaces Val codon by Leu codon"
FT	/*tag= d	insert(51827..51828, cc)
FT	/*tag= d	/note= "insertion, results in frameshift"
XX	W09612033-A1.	
PN	25-APR-1996.	
PD	11-OCT-1995;	95WO-US13357.
XX	31-JAN-1995;	95US-0381520.
XX	12-OCT-1994;	94US-0323443.
XX	(IGIG-) IG LAB INC.	
PA	(UYUO) UNIV JOHNS HOPKINS.	
XX		
PI	Burn TC, Connors TD, Dackowski W, Germino G, Klingner KW;	
PI	Landes GM, Qian F;	
XX	WPI: 1996-222017/22.	
DR		
XX		
PT	Isolated human polycystic kidney disease gene and its mutants -	
PT	carriers for treatment of polycystic kidney disease and screening for	
XX		
ES	Claim 1; Fig 1; 65pp; English.	
XX		
CC	The present sequence is that of the normal human PKD1 gene from	
CC	chromosome 16. Mutations in this gene (e.g. transitions,	
CC	transversions, deletions and/or insertions) are associated with	
CC	adult-onset polycystic kidney disease (APKD). The PKD1 locus is	
CC	GC-rich (62.4%). Comparison of this sequence with a previously	
CC	reported partial cDNA sequence revealed differences at three	
CC	locations (see features table). The most significant difference is	
CC	the presence of two additional cytosine residues on the plus-strand	
CC	at position 4366 of the previously reported sequence. The insertion	
CC	results in a frame-shift in the predicted protein coding sequence,	
CC	leading to replacement of 92 C-terminal amino acids with a novel	
CC	12 amino acid C-terminus. The PKD1 gene contains 23 Alu repeats.	
CC	There is a region consisting of 17 tandem copies of a perfect 27 bp	
CC	repeat and two large CT-rich regions.	
XX		
Q0	Sequence 53577 BP; 8495 A; 17681 C; 15785 G; 11616 T; 0 other;	

Query Match	11.8%;	Score 31.2;	DB 17;	Length 53577;
Best Local Similarity	47.4%;	Pred. No. 3.3;		

Matches	93;	Conservative	0;	Mismatches	103;	Indels	0;	Gaps	0;
OY	18	ccaatgctctcggttccaagctctcaagccgaatcaactgtctctatcagaagcg	77						
Db	35270	cccatgccccctccctccctccctcccatccctccctccctccctccctcccatcc	35329						
OY	78	cttgcccccgagatctcgtctcgtctcaacttcaactcttcgatgctttatcac	137						
Db	35330	atccccctctctccctccctctctccctctctctctccctctctccctctcc	35389						
OY	138	aaaactactatcgcgtctccctccgatacatcaactcgtccctcttaaaacctcta	197						
Db	35390	ctctccctctctcccatccatcccccctccatccatcccccctctctctcccatcc	35449						
OY	198	cggatcaaatgctccc	213						
Db	35450	ctctcccatctctc	35465						
RESULT	6								
T94108									
ID	T94108	standard; DNA; 53577	BP.						
XX									
AC	T94108;								
XX									
DT	01-JUN-1998	(first entry)							
XX									
DE									
XX									
Human PKD1 locus between chromosomal markers ATP1 (ATP6C) and D16S84.									
XX									
KW	Human; polycystic kidney disease 1; PKD1; treatment;								
XX	autosomal dominant polycystic kidney disease; APKD; ss.								
OS	Homo sapiens.								
XX									
PN	W09744457-A1.								
XX									
PD	27-NOV-1997.								
XX									
PF	22-MAY-1997;	97WO-US08799.							
XX									
PR	03-JUN-1996;	96US-0658136.							
XX									
PR	24-MAY-1996;	96US-0655360.							
XX									
PA	(GENZ) GENZYME CORP.								
XX									
PI	Burn T, Connors T, Dackowski W, Germino G, Klinger K;								
XX									
PI	Qian F;								
XX									
DR	WPI; 1998-018511/02.								
XX									
XX									
BS	Example 5; Pages 60-89; 257pp; English.								
XX									
CC	The present sequence is the human polycystic kidney disease 1								
XX	(PKD1) locus between chromosomal markers ATP1 (ATP6C) and D16S84.								
CC	The PKD1 gene or polypeptide may be used to treat autosomal								
XX	dominant polycystic kidney disease (APKD), and identify carriers								
CC	of mutant PKD1 genes, i.e. subjects susceptible to APKD. Antibodies								
XX	(Ab) that distinguish between normal and mutant PKD1 sequences can								
CC	also be used in diagnostic tests. Anti-PKD1 Ab may also be used to								
XX	perform subcellular and histochemical localisation studies, and to								
CC	block the function of PKD1. Ab are also useful in rational drug								
XX	design studies to identify and test inhibitors of PKD1. Sense and								
CC	antisense sequences derived from the PKD1 gene may used for								
XX	detection and therapy.								
CC									
Sequence	53577	BP; 8495	A; 17684	C; 15782	G; 11616	T; 0	other;		
Query Match	11.8%;	Score	31.2;	DB	19;	Length	53577;		
Best Local Similarity	47.4%;	Pred.	NO.	3.3;					

Query Match	11.8%;	Score 31.2;	DB 19;	Length 53577;
Best Local Similarity	47.4%;	Pred. NO. 3.3;		

RESID	7
223892	223892
XX	223892 standard; DNA; 36901 BP.
XX	
AC	223892;
XX	
DT	25-JAN-2000 (first entry)
DE	Murine LOBO genomic DNA fragment 2

OS	Mus musculus.	
XX		
PN	W09J50284-A2.	
XX		
PD	07-OCT-1999.	
XX		
XX		
PF	26-MAR-1999;	99WO-EP02055
XX		
PR	27-MAR-1998;	98DE-1013799
XX		

PA (ROSE/) ROSENTHAL A.
XX
PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T,
XX
DR WPI, 1999-601320/51.
XX

Example 3; Page 97-117; 391pp; German.

uns invention describes novel nucleic acids (1: designated LOBO (Long bones)) encoding proteins influencing bone development in mammals. The proteins of the invention reduce and/or inactivate bone extension (i.e. development), with exception of the skull and have osteopathic activity. The nucleic acid molecules, proteins and antibodies can be used in diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods and nucleic acid molecules, etc. are useful for production of transgenic animals, especially a transgenic mouse for the study of diseases associated with bone development, e.g. spondyloepiphyseal dysplasia and achondroplasia. This sequence encodes the murine LOBO protein described in the method of the invention.

Sequence 36901 BP; 9597 A; 9409 C; 8973 G; 8922 T; 0 other;

Query Match	11.7%;	Score 30.8;	DB 20;	Length 36901;
Best Local Similarity	48.8%;	Pred. No. 3.9;		
Matches 83; Conservative	0;	Mismatches 87;	Indels 0;	Gaps 0

RESULT	8
ID	223897/c
XX	223897 standard; DNA; 30886 BP.
AC	223897;
XX	
DT	25-JAN-2000 (first entry)
XX	
DE	Murine LOBO homologue genomic DNA fragment 3.

OS	Mus musculus.	
XX		
PN	W09J950284-A2.	
XX		
PD	07-OCT-1999.	
XX		
PF	26-MAR-1999;	99WO-EP02055
XX		
PR	27-MAR-1998;	98DE-1013799

PA (ROSE/) ROSENTHAL A.
XX
PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T,
XX
DR MPI; 1999-601320/51.

Example 3; Page 189-211, 391pp; German.

This invention describes novel nucleic acids (1; designated LOBO (long bones)) encoding proteins influencing bone development in mammals. The proteins of the invention reduce and/or inactivate bone extension (i.e. development), with exception of the skull and have osteopathic activity. The nucleic acid molecules, proteins and antibodies can be used in diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods and nucleic acid molecules, etc. are useful for production of transgenic animals, especially a transgenic mouse for the study of diseases associated with bone development, e.g. Spondyloepiphyseal dysplasia and achondroplasia. This sequence encodes the murine LOBO protein described in the method of the invention.

sequence 38886 BP; 10087 A; 9920 C; 9356 G; 9523 T; 0 other;

Query Match	11.7%	Score 30.8;	DB 20;	Length 38886;
Best local Similarity	48.8%	Pred. No. 3.9;		
Matches	83;	Conservative	0;	Mismatches 87;
				Indels 0;
				Gaps 0;

[illegible]

XX Claim 4; Page 249; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic
XX part of a Mycobacterium tuberculosis antigen (Ag). Also described
XX are vaccines and fusion protein containing M. tuberculosis Ag's.
XX M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
XX other polypeptides fragments, can be used in pharmaceutical compositions
XX or vaccines to generate a protective or therapeutic immune response to
XX M. tuberculosis and as reagents in skin tests for diagnosis of
XX tuberculosis. Ag can induce proliferation of, or cytokine secretion
XX by, T, B or natural killer cells and/or macrophages in
XX tuberculosis-immune subjects. 219249 to 219460 and Y39083 to Y39225
XX are used in the exemplification of the present invention.

SQ Sequence 1036 BP; 153 A; 303 C; 177 G; 183 T; 220 other;

Query Match 11.2%; Score 29.6; DB 20; Length 1036;
Best Local Similarity 29.7%; Pred. No. 2.5;
Matches 62; Conservative 37; Mismatches 110; Indels 0; Gaps 0;

QY 5 ctggagcaattcccaatggtccgttccaaagctccaagccaatccattgtctc 64
DB 727 ctcksaamcwnccnccctckaaacccckcyksgmtncccaatntwcmwckcnskntntm 786
QY 65 ctatcagaagcgctcgcccgagattcgtctcctcctcaacttaaccttccat 124
DB 787 ctcccaaytnrcnccnccrctccckstscamwtataaaacccwawynkknkcwm 846
QY 125 cgcctttaacaaaactaactatccgtccctccgataccatccatccctct 184
DB 847 awtamaagcwtcnyncnccnckntktamwccckmccckswtwcykscsccmctct 906
QY 185 taaacctccttaagcatcaatgctccc 213
DB 907 mnaagcccccckktynkwmcccttcccccc 935

RESULT 11
ID 219209
XX 219209 standard; CDNA; 1036 BP.

AC 219209;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen cDNA encoding 5' Erdn-8.

KW Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity; ss.

OS Mycobacterium tuberculosis.

PN WO9942118-A2.

PD 26-AUG-1999.

PF 17-FEB-1999; 99WO-US03265.

PR 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

PA (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PT Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;

DR WPI; 1999-527416/44.

PT New polypeptide comprising antigenic portions of M. tuberculosis

XX Claim 4; Page 294-295; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding
XX nucleic acids derived from Mycobacterium tuberculosis. The novel
XX polypeptides are useful for detecting M. tuberculosis infection in a
XX biological sample by detecting antibodies which bind with the
XX polypeptides, and are useful as vaccines for immunizing against
XX M. tuberculosis infection. The new detection methods are needed as
XX current vaccination strategies do not provide 100% immunity.

SQ Sequence 1036 BP; 153 A; 303 C; 177 G; 183 T; 220 other;

Query Match 11.2%; Score 29.6; DB 20; Length 1036;
Best Local Similarity 29.7%; Pred. No. 2.5;
Matches 62; Conservative 37; Mismatches 110; Indels 0; Gaps 0;

QY 5 ctggagcaattcccaatggtccgttccaaagctccaagccaatccattgtctc 64
DB 727 ctcksaamcwnccnccctckaaacccckcyksgmtncccaatntwcmwckcnskntntm 786
QY 65 ctatcagaagcgctcgcccgagattcgtctcctcctcaacttaaccttccat 124
DB 787 ctcccaaytnrcnccnccrctccckstscamwtataaaacccwawynkknkcwm 846
QY 125 cgcctttaacaaaactaactatccgtccctccgataccatccatccctct 184
DB 847 awtamaagcwtcnyncnccnckntktamwccckmccckswtwcykscsccmctct 906
QY 185 taaacctccttaagcatcaatgctccc 213
DB 907 mnaagcccccckktynkwmcccttcccccc 935

RESULT 12
ID 219209/c
XX 219209 standard; DNA; 1664976 BP.

AC 219209;

DT 10-NOV-1998 (first entry)

DE Methanococcus jannaschii circular chromosome.

KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KW genome; autotrophic; extrachromosomal element; identification; ds.

OS Methanococcus jannaschii.

PN WO9807830-A2.

PD 26-FEB-1998.

PF 22-AUG-1997; 97WO-US14900.

PR 22-AUG-1996; 96US-0024428.

PA (GENO-) INST GENOMIC RES.

PA (UNIT) UNIV ILLINOIS FOUND.

PA (UOJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;

DR WPI; 1998-169145/15.

PT Complete genome sequence of methano-genic archaeon, Methanococcus

PT jannaschii - useful in identification of M. jannaschii genome

XX fragment

XX Claim 13; Page 152-585; 614pp; English.

XX The present sequence represents the complete 1.66-megabase pair genome
XX sequence of the Methanococcus jannaschii circular chromosome. The
XX present invention describes M. jannaschii open reading frames from the


```

FT      /*tag= d
FT      /number= 2
FT      3483..8805
FT      intron
FT      /tag= e
FT      /number= 2
FT      8806..9486
FT      exon
FT      /tag= f
FT      /number= 3
FT      9487..10151
FT      intron
FT      /tag= g
FT      /number= 3
FT      10152..10863
FT      exon
FT      /tag= h
FT      /number= 4
FT      10864..11572
FT      intron
FT      /tag= i
FT      /number= 4
FT      11573..12404
FT      exon
FT      /tag= j
FT      /number= 5
FT      12405..12534
FT      intron
FT      /tag= k
FT      /number= 5
FT      12535..12577
FT      exon
FT      /tag= l
FT      /number= 6
FT      WO9830083-A1.
FT      16-JUL-1998.
FT      PD
FT      XX
FT      PE
FT      09-JAN-1998; 98WO-US00615.
FT      XX
FT      PR
FT      10-JAN-1997; 97US-0781734.
FT      XX
FT      PA
FT      (REGC ) UNIV CALIFORNIA.
FT      PI
FT      Meyers B, Michelmore RM, Shen K;
FT      XX
FT      DR
FT      WPI; 1998-398692/34.
FT      P-PSDB; W64193.
FT      XX
FT      PT
FT      New resistance gene nucleic acids - useful to produce disease
PT      resistant plants, e.g. lettuce, Lactuca sativa, and to detect
PT      resistance genes, e.g. to fingerprint cultivars
PT      XX
PT      PS
PT      Claim 9; Page 158-164; 183pp; English.
PT      XX
CC      This is the nucleotide sequence of the RG25 resistance gene,
CC      which was isolated from genomic DNA of lettuce cultivar Diana
CC      following bacterial artificial chromosome insert characterization
CC      and sequencing. The invention relates to families of resistance
CC      genes, termed RG or RLG, from lettuce. RG families RG1-RG5 and RG7
CC      have been identified, each RG family being defined as a group of
CC      polypeptide sequences that have at least 60% amino acid sequence
CC      identity. Individual members of an RG family typically map to the
CC      same genomic locus. RG nucleic acid sequences (see V44188-257) can
CC      be used to confer resistance in plants to a variety of pests,
CC      including viruses, fungi, nematodes, insects and bacteria.
CC      Sequences from within the RG genes can be used to fingerprint
CC      cultivars or germplasm for the presence of desired resistance
CC      genes. Promoters of RG genes can be used to drive heterologous
CC      gene expression under conditions in which RG genes are expressed.
CC      RG polynucleotides can also be used to produce antisense and
CC      ribozyme molecules useful for inhibiting RG activity in plants.
CC      The invention also provides RG proteins (see W64150-98) that are
CC      encoded by these RG genes.
CC      XX
SQ      Sequence 12793 BP; 4204 A; 2042 C; 2321 G; 4181 T; 45 other:

```

Query Match 11.1%; Score 29.2; DB 19; Length 12793;
 Best Local Similarity 57.8%; Pred. No. 8.7;

```

Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 170 tcactgcctccctttaaacccttcagatcaaatgctcccaacagatggaaatg 229
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8972 TCACTCTCTATCTCAACACACCTCCACCTTACTCTCCCAATTAACTTATG 8913
QY 230 aagtcctcaagaagagacatgctcatgat 259
      || || || || || || || || || || || || || || || ||
Db 8912 AGGTTATGAAAGAGCTCATGACACAGAT 8883

RESULT 15
Z16507/c
ID Z16507 standard; cDNA; 1005 BP.
XX
AC Z16507;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:3977.
XX
KW Human: gene: gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PE 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Erkenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Glase K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
PI XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
PT XX
PS Claim 1; Page 1888-1889; 2479pp; English.
PS XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in Z12532 to Z17779. Also described is a
CC method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in Z12532 to Z17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purposes, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2000, 18:14:35 ; Search time 75.1 Seconds
(without alignments)
531.661 Million cell updates/sec

Title: US-09-300-482-569

Perfect score: 264
Sequence: 1 ctgactgcagcaattccca.....gacatgctcatgatttta 264

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/pdata/1/lna/5A.COMB.seq:*
2: /cgn2_6/pdata/1/lna/5B.COMB.seq:*
3: /cgn2_6/pdata/1/lna/5C.COMB.seq:*
4: /cgn2_6/pdata/1/lna/5D.COMB.seq:*
5: /cgn2_6/pdata/1/lna/6.COMB.seq:*
6: /cgn2_6/pdata/1/lna/CTUS.COMB.seq:*
7: /cgn2_6/pdata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.2	19.4	7218	1	US-08-232-463-14 Sequence 14, Appl
2	31.4	11.9	4002	5	US-09-356-952-9 Sequence 9, Appl
3	31.2	11.8	53526	5	US-08-658-136-2 Sequence 2, Appl
4	31.2	11.8	53577	5	US-08-658-136-1 Sequence 1, Appl
5	28.8	10.9	2575	1	US-08-135-511-33 Sequence 33, Appl
6	28.8	10.9	2575	1	US-08-483-852-10 Sequence 10, Appl
7	28.8	10.9	2575	1	US-08-361-458-5 Sequence 5, Appl
8	28.8	10.9	2575	1	US-08-477-953-10 Sequence 10, Appl
9	28.8	10.9	2575	2	US-08-187-453-33 Sequence 33, Appl
10	28.8	10.9	2575	3	US-08-477-952-10 Sequence 10, Appl
11	27.6	10.5	3131	5	US-09-035-648-23 Sequence 23, Appl
12	27	10.2	933	7	5340934-12 Patent No. 5340934
13	27	10.2	2444	3	US-08-821-355A-2 Sequence 2, Appl
14	27	10.2	2444	4	US-09-003-687A-2 Sequence 2, Appl
15	26.8	10.2	1255	2	US-08-518-878B-38 Sequence 38, Appl
16	26.8	10.2	1255	2	US-08-294-522B-38 Sequence 38, Appl
17	26.8	10.2	1255	2	US-08-470-868A-38 Sequence 38, Appl
18	26.8	10.2	1545	1	US-08-400-275-17 Sequence 17, Appl
19	26.8	10.2	1596	5	US-08-807-861A-38 Sequence 38, Appl
20	26.8	10.2	1596	5	US-09-210-681A-38 Sequence 38, Appl
21	26.6	10.2	1596	5	US-08-946-719A-38 Sequence 38, Appl
22	26.6	10.1	564	7	5175695-2 Patent No. 5175095
23	26.6	10.1	1931	1	US-08-530-492-2 Sequence 2, Appl
24	26.4	10.0	325	3	US-08-332-766A-11 Sequence 11, Appl
25	26.4	10.0	848	5	US-09-009-913-338 Sequence 38, Appl
26	26.4	10.0	2246	6	PCT-US91-08525-23 Sequence 23, Appl

27	26.4	10.0	2246	6	PCT-US91-08525-27 Sequence 27, Appl
28	26.4	10.0	2246	6	PCT-US91-08525-33 Sequence 33, Appl
29	26.4	10.0	2246	6	PCT-US93-04384-3 Sequence 3, Appl
30	26.4	10.0	2246	6	PCT-US93-04384-5 Sequence 5, Appl
31	26.4	10.0	3716	6	PCT-US91-07046-1 Sequence 1, Appl
32	26.4	10.0	4359	3	US-08-566-398-23 Sequence 23, Appl
33	26.4	10.0	4365	6	PCT-US91-08525-25 Sequence 25, Appl
34	26.4	10.0	4365	6	PCT-US93-04384-7 Sequence 7, Appl
35	26.4	10.0	4365	6	PCT-US93-04384-15 Sequence 15, Appl
36	26.4	10.0	4598	1	US-07-807-043B-5 Sequence 5, Appl
37	26.4	10.0	4698	1	US-08-299-849B-5 Sequence 5, Appl
38	26.4	10.0	4698	4	US-08-142-368A-5 Sequence 5, Appl
39	26.4	10.0	4698	5	US-08-967-727-5 Sequence 5, Appl
40	26.4	10.0	4757	1	US-07-865-662F-12 Sequence 12, Appl
41	26.4	10.0	6090	3	US-08-566-398-35 Sequence 35, Appl
42	26.4	10.0	6144	3	US-08-566-398-35 Sequence 35, Appl
43	26.4	10.0	35100	2	US-08-306-691B-19 Sequence 19, Appl
44	26.4	10.0	35100	6	PCT-US93-06251-19 Sequence 19, Appl
45	26.2	9.9	1512	5	US-09-142-514-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHETTLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZ9PT.F15
; US-08-232-463-14

Query Match 19.4%; Score 51.2; DB 1; Length 7218;

Best Local Similarity 5.5%; Pred. No. 2,66-08;

Matches 11; Conservative 128; Mismatches 61; Indels 0; Gaps 0;

QY 15 tccccatgcttcgcttcgaagctcgaagcgaacccaactcgaactgcttccttcagaa 74

Db 1251 yyy 1310

QY 75 ggcgtcccccgcgagatctcgtctcctcaactcgaacttccttcacgtctttaa 134

Db 1311 yyy 1370

QY 135 cacaactaactcgtcgtcctccgacatcgaactcgtcgtccttccttcaaacctcc 194

Db 1371 yyy 1430

QY 195 ttagcatcaatgctccca 214

Db 1431 yyygtaaccaatcttcta 1450

RESULT 2

US-09-356-952-9

Sequence 9, Application US/09356952

Patent No. 6117663

GENERAL INFORMATION:

APPLICANT: Borlack-Sjodin, Ann

APPLICANT: Margaret, S. M.

APPLICANT: Bor-Sogil, Dafna

APPLICANT: Cole, Philip

APPLICANT: Kuriyan, John

TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE

FILE REFERENCE: 600-1-228N

CURRENT APPLICATION NUMBER: US/09/356,952

CURRENT FILING DATE: 1999-07-19

EARLIER APPLICATION NUMBER: 60/093,631

NUMBER OF SEQ ID NOS: 14

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 9

LENGTH: 4002

TYPE: DNA

ORGANISM: Homo sapiens

US-09-356-952-9

Query Match

Best Local Similarity 11.9%; Score 31.4; DB 5; Length 4002;

Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 83 ccccgagatctcgtctcctcgaactcgaacttccttcacgtctttaaatacaaac 142

Db 3665 cactcgtatgtttctcaagctacactacatcgaactccccccttggcaaaaaa 3724

QY 143 taactatcgcgtcctccgacatcgaactcgtcgtccttccttcaaacctcc 195

Db 3725 gtgacatgacatgctccttcctcccaaacagaccccttccttcaaacac 3777

RESULT 3

US-08-658-136-2

Sequence 2, Application US/08658136

Patent No. 6071717

GENERAL INFORMATION:

APPLICANT: KLINGER, KATHERINE W

APPLICANT: LANDES, GREGORY M

APPLICANT: BURN, TIMOTHY C

APPLICANT: CONNORS, TIMOTHY D

APPLICANT: DACKOWSKI, WILLIAM

APPLICANT: GERMINO, GREGORY

APPLICANT: QIAN, FENG

TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS: GENZYME CORPORATION

STREET: ONE MOUNTAIN ROAD

CITY: FRAMINGHAM

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 01701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/658,136

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: LASSEN, ELIZABETH

REGISTRATION NUMBER: 31,945

REFERENCE/DOCKET NUMBER: GEN4-17.8

TELECOMMUNICATION INFORMATION:

TELEPHONE: 508-872-5415

TELEFAX: 508-872-8400

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 53526 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-658-136-2

Query Match

Best Local Similarity 11.8%; Score 31.2; DB 5; Length 53526;

Matches 93; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 18 ccaatggttcgcttcgaagctcgaagcgaacatcgaactgcttcctatcagaagcg 77

Db 35274 CCCATGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 35333

QY 78 ctgcgccccgagatctcgtctcctcgaactcgaacttccttcacgtctttaa 137

Db 35334 ATCCGCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 35393

QY 138 aaactaactatcgcgtcctccgacatcgaactcgaacttccttcacgtctt 197

Db 35394 CTCTCCCTCCTCCTCCATCCCTCCCTCCCTCCCTCCCTCCCTCCTCCTCT 35453

QY 198 cggatcaaatgctccc 213

Db 35454 CCTCCCTCCTCCTC 35469

RESULT 4

US-08-658-136-1

Sequence 1, Application US/08658136

Patent No. 6071717

GENERAL INFORMATION:

APPLICANT: KLINGER, KATHERINE W

APPLICANT: LANDES, GREGORY M

APPLICANT: BURN, TIMOTHY C

APPLICANT: CONNORS, TIMOTHY D

APPLICANT: DACKOWSKI, WILLIAM

APPLICANT: GERMINO, GREGORY

APPLICANT: QIAN, FENG

TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS:

ADDRESS: GENZYME CORPORATION

STREET: ONE MOUNTAIN ROAD

```

? FILING DATE : 13-OCT-1993
? CLASSIFICATION : 435
? ATTORNEY/AGENT INFORMATION :
? NAME : SANDERCOCK, COLIN G.
? REGISTRATION NUMBER : 31,298
? REFERENCE/DOCKET NUMBER : 18748/175
? TELECOMMUNICATION INFORMATION :
? TELEPHONE : (202)672-5300
? TELEFAX : (202)672-5399
?
? TELEX : 904136
?
? INFORMATION FOR SEQ ID NO: 33:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2575 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? MOLECULE TYPE: DNA (genomic)
?
? US-08-135-511-33

```

Query Match	10.9%;	Score 28.8;	DB 1;	length 2575;
Best Local Similarity	52.5%;	Pred. NO. 1.2;		
Matches 63;	Conservative 0;	Mismatches 57;	Indels 0;	Gaps 0;

OY	124	tcgcctttaacaaacaaactcgcgtccctcgatcaacaattacgtccctc	183
Db	1010	TCACATCAGAGAAAAAGCTCCACATCTTCTCTCTCTCCCTTTCTCTCCCTCCC	1065
OY	184	ttaaacgcctcttaagatcaaatgctcccaacagatgaggaatgcctgaaga	243
Db	1070	TGACTGCCCTCTCTCTATCTCTCTCTCTCTCCCTGACCTGGCAAGGTATATGGTGTGCAG	1129

RESULT 6
 US-08-483-852-10
 : Sequence 10, Application US/08483852
 : Patent No. 5650286
 :
 : GENERAL INFORMATION:
 : APPLICANT: CHIANG, John Young Ling
 : TITLE OF INVENTION: Genomic DNA of Human Cholesterol
 : TITLE OF INVENTION: 7a-Hydroxylase and Methods of Using It
 : NUMBER OF SEQUENCES: 11
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Foley & Lardner
 : STREET: 3000 K Street, N.W., Suite 500
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20007-5109
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/483,852
 : FILING DATE: 07-JUN-1995
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/361,458
 : FILING DATE: 21-DEC-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/135,511
 : FILING DATE: 13-OCT-1993
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/135,488
 : FILING DATE: 13-OCT-1993
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/135,510
 : FILING DATE: 13-OCT-1993
 : ATTORNEY/AGENT INFORMATION:
 : NAME: SANDERCOCK, Colin G.
 : REGISTRATION NUMBER: 31,298
 : REFERENCE/DOCKET NUMBER: 18748/220 HOCE

[illegible]

RESULT 8
 US-08-477-953-10
 ; Sequence 10, Application US/08477953
 ; Patent No. 5677159
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: CHIANG, John Young Ling
 ; TITLE OF INVENTION: Genomic DNA of Human Cholesterol
 ; TITLE OF INVENTION: 7 α -Hydroxylase and Methods of Using It
 ; NUMBER OF SEQUENCES: 11
 ;
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.

```

1 ZIP: 20007-5109
2 COMPUTER READABLE FORM:
3 MEDIUM TYPE: Floppy disk
4 COMPUTER: IBM PC compatible
5 OPERATING SYSTEM: PC-DOS/MS-DOS
6 SOFTWARE: Patentin Release #1.0, Version #1.25
7 CURRENT APPLICATION DATA:
8 APPLICATION NUMBER: US/08/477,953
9 FILING DATE: 07-JUN-1995
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: US 08/361,458
12 FILING DATE: 21-DEC-1994
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: US 08/135,511
15 FILING DATE: 13-OCT-1993
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US 08/135,488
18 FILING DATE: 13-OCT-1993
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US 08/135,510
21 FILING DATE: 13-OCT-1993
22 ATTORNEY/AGENT INFORMATION:
23 NAME: SANDERCOCK, Colin G.
24 REGISTRATION NUMBER: 31,298
25 REFERENCE/DOCKET NUMBER: 18748/219 HOCE
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (202)672-5300
28 TELEFAX: (202)672-5399
29 TELEX: 904136
30 INFORMATION FOR SEQ ID NO: 10:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 2575 base pairs
33 TYPE: nucleic acid
34 STRANDEDNESS: single
35 TOPOLOGY: linear
36 MOLECULE TYPE: DNA (genomic)
37 US-08-477-953-10
38
39 Query Match 10.9%; Score 28.8; DB 1; Length 2575;
40 Best Local Similarity 52.5%; Pred. No. 1.2;
41 Matches 63; Conservative 0; Mismatches 57; Indels 0; Gaps 0.
42
43 124 tgcgtttatcacaaactcactccctccctccgatccatctacatccgtccctc 183

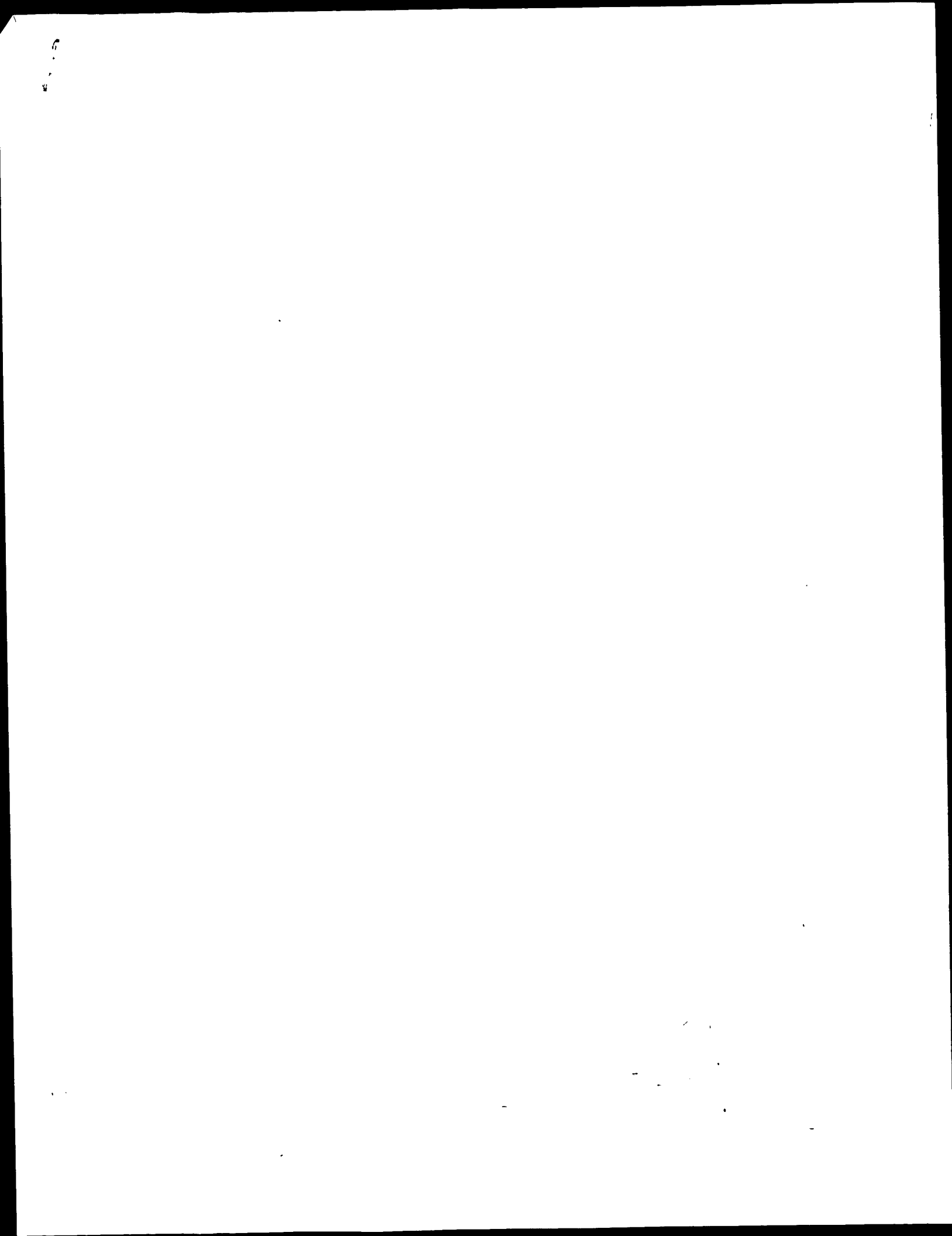
```

Journal of Management Inquiry 23(4) 409–427 © The Author(s) 2014. Reprints and permissions: sagepub.com/journalsPermissions.nav


```

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: Pennie & Edmonds
3 STREET: 1155 Avenue Of The Americas
4 CITY: New York
5 STATE: New York
6 COUNTRY: U.S.A.
7 ZIP: 10036-2711
8
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: PatentIn Release #1.0, Version #1.30
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/08/518,878B
16 FILING DATE: 23-AUG-1995
17 CLASSIFICATION: 435
18
19 ATTORNEY/AGENT INFORMATION:
20 NAME: Coruzzi, Laura A.
21 REGISTRATION NUMBER: 30,742
22 REFERENCE/DOCKET NUMBER: 7853-036
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (212) 790-9090
25 TELEFAX: (212) 869-9741/8864
26 TELEX: 66141 PENNIE
27
28 INFORMATION FOR SEO ID NO: 38:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 1255 base pairs
31 TYPE: nucleic acid
32 STRANDEDNESS: single
33 TOPOLOGY: linear
34
35 US-08-518-878B-38

```



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2000, 18:13:15 ; Search time 823.97 Seconds
(without alignments)
1980.971 Million cell updates/sec

Title: US-09-300-482-569

Perfect score: 264
Sequence: 1 ctgactgcagcaatcccca.....gacatgcttcgatcttta 264

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estfun:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estp11:*
71: em_estp12:*
72: em_estp13:*
73: em_estp14:*
74: em_estp15:*
75: em_estp16:*
76: em_estro1:*
77: em_estro2:*
78: em_estro3:*
79: em_estro4:*
80: em_estro5:*
81: em_estro6:*
82: em_estro7:*
83: em_estro8:*
84: em_estro9:*
85: em_estro10:*
86: em_estro11:*
87: em_estro12:*
88: em_estro13:*
89: gb_gss1:*
90: gb_gss2:*
91: gb_gss3:*
92: gb_gss4:*
93: em_gss1:*
94: em_gss2:*
95: em_gss3:*
96: em_gss4:*
97: gb_gss5:*
98: gb_gss6:*
99: gb_gss7:*
100: gb_gss8:*
101: gb_gss9:*
102: em_gss5:*
103: em_gss6:*
104: em_gss7:*
105: em_gss8:*
106: em_gss9:*
107: em_gss10:*
108: gb_gss11:*
109: gb_gss12:*
110: em_gss13:*
111: gb_gss14:*
112: gb_gss15:*
113: gb_gss16:*
114: gb_gss17:*
115: gb_gss18:*
116: gb_gss19:*

ORIGIN

Query Match 95.5%; Score 252; DB 20; Length 540;
 Best Local Similarity 99.6%; Pred. No. 9.3e-68;
 Matches 263; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

2 tgactgcagcaatcccaatgcttcgcttcgaagctcccaagcgaacatccacttcg 61
 |||||||
 Db 3 TGACTTGAGCAATTCCTCCATGGCTTCGCTTCAGAGCTCCCAAGCCCAATCCACTTGC 62
 Qy 62 ttccctatcagaagcgttcgccccgagatctcgtctccatccatccacttcctc 121
 |||||||
 Db 63 TTCCTTATCAGAAAGCGCTTCGCCGCCGAGATTCTCGCTTCCTCCACTTCAACCTTCTTC 122
 Qy 122 catgccttatacaaaaactaatcctcgtctccctccatccatccacttcctcc 181
 |||||||
 Db 123 CATGCGCTTTATATCAAAACTAATCACTATCCGCTTCCTCCGATCCATCATCTCGCTCC 182
 Qy 182 tcttaaacctccttaagcaatgctcccaacagatgggaatggaatcctgcaaa 241
 |||||||
 Db 183 TCTTAAACCTCCTTACGATCAATGCTCCCAACAGATGGAGATGGAATGGAATGCTCGCAA 242
 Qy 242 gagaca-tgcttcagtacttta 264
 |||||||
 Db 243 GAGGACAGTGTCTCATGATCTTTA 266

RESULT 2
 AM569249 547 bp mRNA EST 18-JUL-2000
 LOCUS s164901.y1 Gm-r1030 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-r1030-3721 5' similar to TR:004894 004894 TRANSLADOLASE ;, mRNA
 sequence.
 ACCESSION AM569249.1 GI:7233907
 VERSION AM569249.1
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 547)
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
 'A., Bolla, B., Marra, M., Hillier, L., Kuabara, T., Martin, J., Beck, C.,
 'Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 'R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 'R., Waterston, R. and Wilson, R.

TITLE
 JOURNAL
 COMMENT
 Public Soybean EST Project
 Contact: Shoemaker R./Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

FEATURES
 source
 location/Qualifiers
 1..547
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-r1030-3721"
 /clone_id="Gm-r1030"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; site_1: SalI; site_2: NotI; This
 High quality sequence stop: 438.
 infoGenomesystems.com web site: www.genomesystems.com
 427-3324 or contact: clones@genomesystems.com or
 call: (800) 430-0030 or (314) 427-3322 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 Inert Length: 1708 Std Error: 0.00

CDNA library was constructed from mRNA isolated from
 immature cotyledons of greenhouse grown plants
 (individual seed fresh weight of 100-300mg). The library
 was prepared using the Life Technologies pSPORT1 cDNA
 library construction kit. Complementary DNA was
 synthesized from mRNA using a poly(dT) sequence with a
 NotI restriction site. SalI linkers adapters were ligated
 to the blunt-ended cDNA fragments followed by NotI
 digestion. The cDNA fragments were directionally cloned
 into the NotI-SalI restriction site of the pSPORT1
 vector. The ligated cDNA fragments were transformed into
 E. coli Electromax DH10B host cells. This library was
 constructed by Dr. Lila Vodkin and Dr. Ann Khanna. Note
 that Gm-r1030 is a re-track of Gm-cl007."

Query Match 94.8%; Score 250.4; DB 23; Length 547;
 Best Local Similarity 99.2%; Pred. No. 2.9e-67;
 Matches 262; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

2 tgactgcagcaatcccaatgcttcgcttcgaagctcccaagcgaacatccacttcg 61
 |||||||
 Db 10 TGACTTGAGCAATTCCTCCATGGCTTCGCTTCAGAGCTCCCAAGCCCAATCCACTTGC 69
 Qy 62 ttccctatcagaagcgttcgccccgagatctcgtctccatccatccacttcctc 121
 |||||||
 Db 70 TTCCTTATCAGAAAGCGCTTCGCCGCCGAGATTCTCGCTTCCTCCACTTCAACCTTCTTC 129
 Qy 122 catgccttatacaaaaactaatcctcgtctccctccatccatccacttcctcc 181
 |||||||
 Db 130 CATGCGCTTTATATCAAAACTAATCACTATCCGCTTCCTCCGATCCATCATCTCGCTCC 189
 Qy 182 tcttaaacctccttaagcaatgctcccaacagatgggaatggaatcctgcaaa 241
 |||||||
 Db 190 TCTTAAACCTCCTTACGATCAATGCTCCCAACAGATGGAGATGGAATGGAATGCTCGCAA 249
 Qy 242 gagaca-tgcttcagtacttta 264
 |||||||
 Db 250 GAGGACAGTGTCTCATGATCTTTA 273

RESULT 3
 AM569131 565 bp mRNA EST 18-JUL-2000
 LOCUS s163901.y1 Gm-r1030 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-r1030-3625 5' similar to TR:004894 004894 TRANSLADOLASE ;, mRNA
 sequence.
 ACCESSION AM569131.1 GI:7233788
 VERSION AM569131.1
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 565)
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
 'A., Bolla, B., Marra, M., Hillier, L., Kuabara, T., Martin, J., Beck, C.,
 'Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 'R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 'R., Waterston, R. and Wilson, R.

TITLE
 JOURNAL
 COMMENT
 Public Soybean EST Project
 Contact: Shoemaker R./Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert Length: 1744 Std Error: 0.00
High quality sequence stop: 415.
Location/Qualifiers

FEATURES

source

1. 565
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-r1030-3625"
/clone_id="Gm-r1030"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site: 1: SalI; Site: 2: NotI; This cDNA library was constructed from mRNA isolated from immature cotyledons of greenhouse grown plants (individual seed fresh weight of 100-300mg). The library was prepared using the Life Technologies pSPORT1 cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restriction site. SalI linkers followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E. coli Electromax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Ann Khanna. Note that Gm-r1030 is a re-track of Gm-cl007."
BASE COUNT 154 a 145 c 112 g 153 t 1 others
ORIGIN

Query Match 94.8%; Score 250.4; DB 23; Length 565;
Best Local Similarity 99.2%; Pred. No. 2.9e-67;
Matches 262; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 2 tgaattgagcaatcccaatggtcttcctcaagatcctcaagcccaatccacttgc 61
DB 10 TGACTTGCACCAATCCCAATGCGCTTCCTTAAAGCTCTCAAGCCCAATCCACTTGC 69
QY 62 ttcttcaagaagcgttcgccccgagattctgccttcctcaacttcaaacctcttca 121
DB 70 TTCTTTATCAGAAAGCGTTTGGCCCGAGATTTCTCGTTCTTCACTTCAACCTTCTTC 129
QY 122 catgcttttaatacaaaactaactatcgcgttccttcgcatcattcctgcctcc 181
DB 130 CATGCTTTTAATCAAAAACATACTATCCGCTTCCTCCGATTCATTCACCTGCTCC 189
QY 182 tcttaaaacctcttaagataaattgtcccaaacagatgggaatggaatctgcaaa 241
DB 190 TCTTAAACCTCTTACGATCAATATGCTCCCAACAGATGGATGGAAAGTCTGCAAA 249
QY 242 gagagaca-tgctcatgatcttta 264
DB 250 GAGGACAGTGGCTTCATGATCTTTA 273

RESULT 4
AM395473 510 bp mRNA EST 17-JUL-2000
LOCUS sn49f02.y1 Gm-cl017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl017-5044 5' similar to TR:004894 004894 TRANSALDOLASE ;, mRNA
sequence.
ACCESSION AM395473
VERSION AM395473.1 GI:6913943
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 510)

AUTHORS

Shoemaker R., Keim P., Vodkin L., Erpelting J., Corryell V., Khanna A., Bolla B., Marra M., Hillier L., Kueba T., Martin J., Beck C., Wylie T., Underwood R., Steptoe M., Rhelsing B., Allen M., Bowers Y., Person B., Swaller T., Gibbons M., Page D., Harvey N., Schurk R., Rittler E., Kohn S., Shin T., Jackson Y., Cardenas M., McCann R., Waterston R. and Wilson R.
Public Soybean EST Project
Unpublished (1999)

JOURNAL

Contact: Shoemaker R./Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

COMMENT

Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert Length: 1816 Std Error: 0.00
Seq primer: -40RP from Glibco
High quality sequence stop: 395.
Location/Qualifiers

FEATURES

source

1. 510
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl017-5044"
/clone_id="Gm-cl017"
/tissue_type="vegetable buds of field grown plants"
/lab_host="XL10-Gold"
/note="Vector: Bluescript II XR; Site: 1: EcoRI; Site: 2: XhoI; This cDNA library was constructed from mRNA isolated from vegetable buds of field grown plants. The cDNA library was prepared using the Stratagene Bluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."
BASE COUNT 141 a 138 c 98 g 133 t
ORIGIN

Query Match 91.7%; Score 242; DB 21; Length 510;
Best Local Similarity 99.6%; Pred. No. 1.2e-64;
Matches 253; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 12 caattcccaatggtcttcggttcaagatctcaagcaaaatccactggtctctatca 71
DB 1 CAATTCCCAATGCGTTTCGTTTCAAGCTCTCAAGCCCAATCCACTGTTCTTATCA 60
QY 72 gaagcgttcgccccgagatctcgtcttcctcaacttcaaaccttcttcacatgcttt 131
DB 61 GAAGCGCTTTCGCCCCGAGATTTCTGCTTCACCTTCAAACTTCTTCAATGCTTTT 120
QY 132 aatcaaaaaaactaactatcgcgttccttcgcatcattcactgctctccctttaaacc 191
DB 121 AATCACAAAAACATACTATCCGCTTCCTCCGATTCATTCAGTCCCTCTTAAAC 180
QY 192 tcttaagataaattgtcccaaacagatgggaatggaatcctgcaaaagagaca-tg 250
DB 181 TCCTTAGGATCAAAATGCTCCCAACAGATGGGATGGAAAGTCTGCAAAAGGACATG 240
QY 251 ctctcatgatcttta 264
DB 241 CTTCATGATCTTTA 254

Email: estlewalson@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com OR
info@genomesystems.com web site: www.genomesystems.com
Possible reversed clone: Similarly on wrong strand
Insert Length: 1680 Gld Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 353
POLYA-NO.

Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax: 520-523-7500, email: paul.keim@na.u.edu, virginia.coryell@na.u.edu"

BASE COUNT 116 a 126 c 79 g 114 t

ORIGIN

Query Match 87.1%; Score 230; DB 10; Length 435;
Best Local Similarity 99.6%; Prid. No. 5,8e-61;
Matches 241; Conservative 0; Mismatches 0; Indels 1; Gaps 1

QY 24 gccctcggttcgaagcctctcaacgccaaatccacttgcttccattataagaagcgttcg 83
|||||
Db 1 gccttcggtttccaacctctctcaagcccAAATCCATTGCTTATCAGAGCGCTTCGC 60
|||||

QY 84 ccccgagatctgcgtctcctcaccttaaacctctctcatcgctttaacaanaact 143
|||||
Db 61 CCCCGAGATCTCGCTTCTCACACTTCAAAACCTTTTCATTCGCTTTTAATCACAANA 120
|||||

QY 144 aaacttcgcgttcctccctcgatccattcactgcgtccctcttaaactccttacgac 203
|||||
Db 121 AACCTATCCGCTTCCCTCCGATCATTCACCTCGCTCCTCTTAAAACCTCTTAGGATC 180
|||||

QY 204 aaatgctccccaaacagatggaaatggaaatgcctcgcaaaaggagca-tgcttatgatct 262
|||||

Db 181 AATGTCTCCCAACACATGGAGATGGAAATGTCAGCAAAGAGACAGTCTTCATGATCTT 240
|||||

QY 263 ta 264
||

Db 241 TA 242

RESULT 6
BE475226

LOCUS BE475226 532 bp mRNA EST 28-JUL-2000

DEFINITION sp72904.y1 Gm-cl044 Glycine max cDNA clone GENOME SYSTEMS CLONE ID Gm-cl044-1615 5' similar to TR-004894 004894 TRANSALDOLASE,, mRNA sequence.

ACCESSION BE475226

VERSION BE475226.1

KEYWORDS GI:9565717

SOURCE soybean.
ORGANISM Glycine max
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 532)

REFERENCE Shoemaker,R., Keim,P., Vodkin,L., Erpelind,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.C., Wylie,T., Underwood,K., Steptoe,M., Theisinger,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pope,D., Harvey,N., Schurk,R., Ratter,E., Koln,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Walters,D., and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST project
Public Soybean EST project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel.: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

TITLE This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 OR (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com High quality sequence stop: 360.

JOURNAL COMMENT

FEATURES

source

1..532

/organism="Glycine max"

```

/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1044-1615"
/clone_1ib="Gm-c1044"
/tissue_type="Hypocotyl, 9-10 day old etiolated seedlings"
/lab_host="DH10B"
/notes="Vector: plasmidscriptII SK+, Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from etiolated hypocotyl tissue of 9-10 day old seedlings of the cultivar Williams. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) primer with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by digestion with EcoRI and XhoI. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the plasmidscript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker."

```

This clone is available through: Genome Systems, Inc., 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3232 FAX: (888) 919-3334 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert length: 1881
High quality sequence stop: 422.
Location/Qualifiers
1..630

Query Match	75.9%	Score 200.4	DB 35	Length 532
Best Local Similarity	92.9%	Pred. No. 9.3e-52		
Matches 221	Conservative	0	Mismatches 16	Indels 1
Oy	28	ccgttccaaagctctcaacagccaatccacttgctcttcctatcagaagcgtctgcgcc	87	
Db	6	ccattttcaagctctcaacagccaatccacttgctcttcctatcagaagcgtctgcgcc	65	
Oy	88	gagattctgcgtcttccacacitcaaaccttcttcacatgctcttatacaaaactaac	147	
Db	66	gagattttcgcgtcttccacacitcaaaccttcttcacatgctcttatacaaaactaac	125	
Oy	148	tatcgcgtcttccctccgatccatcactcgtctccctcttaaacctccatagatcaat	207	
Db	126	tctccgcttccctccgatccatcactcgtctccctcttaaacctccatagatcaat	185	
Oy	208	gctcccaacagatggaatggaatgcctctgcaagagaca-tgcttcatactctta	264	
Db	186	gctcccaacagatggaatggaatgcctctgcaagagaca-tgcttcatactctta	243	
RESULT	7			
LOCUS	AM201037	690 bp	mRNA	EST
DEFINITION	se97b07.y1 Gm-cl027 Glycine max cDNA clone GENOME SYSTEMS CLONE ID Gm-cl027-686 5' similar to IR:004894 004894 TRANSALDOLASE ; mRNA sequence.			
ACCESSION	AM201037			
VERSION	AM201037.1	GI:6481766		
KEYWORDS	EST.			
SOURCE	Soybean.			
ORGANISM	Glycine max			
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta			
AUTHORS	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.			
	1 (bases 1 to 690)			
	Shoemaker, R., Kelm, P., Vodtin, L., Eprelding, J., Cortell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritzer, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.			
	Public Soybean EST Project			
	Unpublished (1999)			
TITLE	Contact: Shoemaker R/Public Soybean EST Project			
JOURNAL	Public Soybean EST Project			
COMMENT	Washington University School of Medicine			
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: est@watson.wustl.edu			

BASE COUNT	ORIGIN
189 a	161 c 145 g 192 t 3 others

Query Match	75.5%;	Score 199.4;	DB 20;	Length 690;
Best Local Similarity	92.8%;	Pred. No. 2e-81;		
Matches 220;	Conservative	0;	Mismatches 16;	Indels 1;
				Gaps
QY	29	cglttccaagctctcaagcgaatccacttgcctctctatcagaagcgtctggcccg	88	
Db	6	CATTTCACCTCTCAACGGCATCTCCACTCTTCCTTATCAGAGGGCTGGCCCAAG	65	
QY	89	agatttcggttcctcaacctcaaacctctctccatcgctcttatacacaanaact	148	
Db	66	AGATTTCGGTCTCTCTCTTCAATCCTTTTCGAACGGTATTATTCAGAAATTAACCT	125	
QY	149	atccgcttcctccgcagatcattcaatcgtctccctcttaaacctccttaagatcaaatg	208	
Db	126	CTCCGCTCTCCCTCGAATCCATTCACCTCCCTCTTAAACCTTCCTTGGCGATCAATG	185	
QY	209	cttcccaaacgattgggaatgaagctctgcaagaaggaca-tgcctcatgactctta	264	
Db	186	CTCCCAACCAATGAGATGAAATGAAGCTCTCGAAGAGACAGTGCCTCATGATCTTTA	242	

RESULT 8

BE211683
LOCUS 485 bp mRNA EST 29-JUN-2000
DEFINITION s064h11.y1 Gm-c1040 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1040-478 5' similar to TR:004894 004894 TRANSALDOLASE ; mRNA sequence.
ACCESSION BE211683
VERSION BE211683.1 GI:8827953
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 485)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 280.
FEATURES
source
location/Qualifiers
1..485
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1040-478"
/clone_lib="Gm-c1040"
/tissue_type="Hypocotyl and Plumule, germinating seeds"
/lab_host="DH10B"
/note="Vector: p773Pac (Pharmacia); Site: 1: EcoRI; Site: 2: NotI. This cDNA library was constructed from mRNA isolated from hypocotyl and plumule tissues of seeds germinated for three days of the cultivar Williams 82. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a NotI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by digestions with EcoRI and NotI. The cDNA fragments were directionally cloned into the EcoRI-NotI restriction site of the p773-Pac vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker."

BASE COUNT 128 a 129 c 102 g 125 t 1 others
ORIGIN

Query Match 75.4%; Score 199; DB 34; Length 485;
Best Local Similarity 93.2%; Pred. No. 2.5e-51;
Matches 219; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 31 ttccaagctcacaacgaacacactgtctctcattacagaagcgttcgccccgag 90
DB 1 ttttcaagctcacaacgaacacactgtctctcattacagaagcgttcgccccgag 60
QY 91 attcgtctcctcaccctcaaacctcttcacgtcttatacaaacactaactat 150
DB 61 atttctgcttctcctcctccttcaaacctcttatacaaacactaactat 120
QY 151 ccgcttcctcctcgtatcattcactcgtctcctccttaaacctccttaagatgaatgct 210

DB 121 cccgcttcctcctcgtatcattcactcgtctccttcaaacctccttgcgattcaatgct 180
QY 211 cccaacagatggagatggagatcctgcgaagagaca-tgcttcattcattta 264
DB 181 cccaacagatggagatggagatcctgcgaagagaca-gagacagcttcattcattta 235

RESULT 9
LOCUS AM132188
DEFINITION s067c11.y1 Gm-c1008 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1008-1389 5' similar to TR:004894 004894 TRANSALDOLASE ; mRNA sequence.
ACCESSION AM132188
VERSION AM132188.1 GI:6133795
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 519)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert Length: 1732 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 449.
FEATURES
source
location/Qualifiers
1..519
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1008-1389"
/clone_lib="Gm-c1008"
/lab_host="DH10B"
/note="Vector: pSPOR1; Site: 1: SalI; Site: 2: NotI. This cDNA library was constructed from mRNA isolated from whole young pods, approximately 2cm long, of 12-week-old greenhouse grown plants. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a NotI restriction site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPOR1 vector. The ligated cDNA fragments were transformed into E. coli ElectroMax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."

JOURNAL COMMENT
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert Length: 1732 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 449.
FEATURES
source
location/Qualifiers
1..519
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1008-1389"
/clone_lib="Gm-c1008"
/lab_host="DH10B"
/note="Vector: pSPOR1; Site: 1: SalI; Site: 2: NotI. This cDNA library was constructed from mRNA isolated from whole young pods, approximately 2cm long, of 12-week-old greenhouse grown plants. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a NotI restriction site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPOR1 vector. The ligated cDNA fragments were transformed into E. coli ElectroMax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."

Query Match 75.3%; Score 198.8; DB 20; Length 519;
Best Local Similarity 92.5%; Pred. No. 2.9e-51;

BASE COUNT 141 a 135 c 101 g 142 t
ORIGIN

Matches 209; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 22 tggcttcgttcacagctcacaacacacactgcttccttaccagaagcgttc 81
 Db 1 tggccctccattttcagctcctcagccatctccagctgcttccatcagagcgcttc 60
 QY 82 gccccgagatctcgtcttcctcacttaaaccttcttcacgtcttcaacacaa 141
 Db 61 gccccagagattctgccttcctcctcctcctcctccttcaaaccttatttatttcaaaa 120
 QY 142 ctaactatccgcttcctcctcctcctcctcctcctccttcaaaccttcaagga 201
 Db 121 ttaacctctccgcttcctcctcctcctcctcctcctttaaactcttgcgga 180
 QY 202 tcaaatgctcccaacagatgggaatgagatcctgcaagaagagac 247
 Db 181 tcaaatgctcccaacagatgggaatgagatcctgcaagaagagac 226

RESULT 10

A1748238

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Glycine max

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;

Fabales; Fabaceae; Papilionoideae; Glycine.

1 (bases 1 to 529)

Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna

A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,

Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers

Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk

R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann

R., Waterston, R. and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World

Parkway Circle St. Louis, Missouri 63134 For further information

call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)

427-3324 or contact: clones@genomesystems.com or

info@genomesystems.com web site: www.genomesystems.com

Possible reversed clone: similarity on wrong strand

Insert length: 1638 Std Error: 0.00

High quality sequence stop: 414.

Location/Qualifiers

1..529

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl011-280"

/issue_type="Immature cotyledons of greenhouse grown

plants"

/lab_host="DH10B"

/note="Vector: Bluescript II SK+; Site_1: EcoRI; Site_2:

XhoI; This cDNA library was constructed from mRNA isolated

from immature cotyledons (100-200mg) of greenhouse grown

plants. The cDNA library was prepared using the ligo

Technologies psupertscript cDNA library construction kit.

Complementary DNA was synthesized from mRNA using a poly

(dT) sequence with a Not I restrictions site. Sal I

location/Qualifiers

1..363

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl011-280"

/issue_type="Immature cotyledons of greenhouse grown

plants"

/lab_host="DH10B"

/note="Vector: Bluescript II SK+; Site_1: EcoRI; Site_2:

XhoI; This cDNA library was constructed from mRNA isolated

from immature cotyledons (100-200mg) of greenhouse grown

plants. The cDNA library was prepared using the ligo

Technologies psupertscript cDNA library construction kit.

Complementary DNA was synthesized from mRNA using a poly

BASE COUNT

ORIGIN

145 a 133 c 104 g 146 t 1 others

Query Match

Best Local Similarity 93.5%; Pred. No. 1.4e-50;

Matches 216; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

linkers adapters were ligated to the blunt-ended cDNA

fragments followed by NotI digestion. The cDNA fragments

were directionally cloned into the NotI-Sal I restriction

site of the pSPORT 1 vector. The ligated cDNA fragments

were transformed into E. coli ElectroMax DH10B host cells.

This library was constructed by Dr. Lala Vodkin and Dr.

Anu Khanna.

QY 35 caagctcacaagcacaatcactgcttccttaccagaagcgttcgcccgaattc 94

Db 7 CAAGCTCAACGCGCATCTCCAGCTGCTCTTATCAGAGCGCTTCGCCAGAGATTC 66

QY 95 tgccttcacactcacaacacttctcactcgtcttcaacacaaactaacctacgcg 154

Db 67 TGCCTCTCTCTCTTCATCTCTTCTTCCACGCTATTAATTACAAATTAACCTCTCCG 126

QY 155 ttcctcgcacatcactcactgcttccttcaaaccttcaagcgaatgctccca 214

Db 127 TTCCCTCGATCCATTCACCTCCCTCTTAAACCTCTTGGGATCAAAATGCTCCCA 186

QY 215 aacgagtggaatggaatgctcctgcaagaagagac-tgcttcacatctta 264

Db 187 AACAGATGGAGATGGAATGCTCTGCAAGAGACAGTCTTCATATCTTTA 237

RESULT 11

A1441149

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Glycine max

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;

Fabales; Fabaceae; Papilionoideae; Glycine.

1 (bases 1 to 363)

Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna

A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,

Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers

Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk

R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann

R., Waterston, R. and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World

Parkway Circle St. Louis, Missouri 63134 For further information

call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)

427-3324 or contact: clones@genomesystems.com or

info@genomesystems.com web site: www.genomesystems.com

Insert length: 1700 Std Error: 0.00

High quality sequence stop: 346

POLYA-No.

Location/Qualifiers

1..363

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl011-280"

/issue_type="Immature cotyledons of greenhouse grown

plants"

/lab_host="DH10B"

/note="Vector: Bluescript II SK+; Site_1: EcoRI; Site_2:

XhoI; This cDNA library was constructed from mRNA isolated

from immature cotyledons (100-200mg) of greenhouse grown

plants. The cDNA library was prepared using the ligo

Technologies psupertscript cDNA library construction kit.

Complementary DNA was synthesized from mRNA using a poly

(dT) sequence with a Not I restrictions site. Sal I

location/Qualifiers

1..363

/organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1004-3586"
 /clone_1lb="Gm-c1004"
 /tissue_type="root"
 /lab_host="Xcl10-Gold"
 /note="Vector: pBluescript II Xr; Site 1: EcoRI; Site 2: XhoI; Root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. Stratagene's first-strand synthesis primer was used [GAGAGAGAGAGAGAGAGACTGCTCGAC(T)-18]. After second-strand synthesis, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected with their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GIBCOBRL life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratagene's pBluescript II Xr Predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies 9n-15) have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax: 520-523-7500, email: paul.keim@nau.edu, virginia.coryell@nau.edu"

BASE COUNT 94 a 108 c 59 g 100 t 2 others
 ORIGIN

Query Match 70.2%; Score 185.4; DB 10; Length 363;
 Best Local Similarity 92.0%; Pred. No. 3.8e-47;
 Matches 206; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 42 tcaagcgaatacctctgcttctatcagaagcgcttcgcccgaatctcgttc 101
 Db 1 tcaagcgaatacctctgcttctatcagaagcgcttcgcccgaatctcgttc 60
 QY 102 ctacactcaaacctctcctcgttttaatacaaaactaatcgttcctc 161
 Db 61 ctctcttcaatctcttccacagctatttaatacaaaactaatcgttcctc 120
 QY 162 cgaatcattcactcgtctcctttaaactcctcgaatcgttcgaatcgttc 221
 Db 121 cgaatcattcactcgtctcctttaaactcctcgaatcgttcgaatcgttc 180
 QY 222 ggggaatgggaagctctgcaagagaca-tgcttcattgatcttta 264
 Db 181 ggggaatgggaagctctgcaagagaca-tgcttcattgatcttta 224

RESULT 12
 AMB30435 452 bp mRNA EST 18-JUL-2000
 LOCUS sm26909.y1 Gm-c1028 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1028-5081 5' similar to TR:004894 004894 TRANSALDOLASE ;, mRNA sequence.
 ACCESSION AMB30435
 VERSION AMB30435.1 GI:7924409
 KEYWORDS EST
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;

REFERENCE

1 (bases 1 to 452)
 Shoemaker R., Keim P., Vodkin L., Erpelting J., Coryell V., Khanna A., Bolla B., Marra M., Hillier L., Kuabada T., Martin J., Beck C., Wylie T., Underwood K., Steptoe M., Theisinger B., Allen M., Bowers J., Peterson B., Swaller T., Gibbons M., Pape D., Harvey N., Schurk R., Rutter E., Kohn S., Shin T., Jackson Y., Cardenas M., McCann R., Waterson R. and Wilson R.
 Public Soybean EST Project
 Unpublished (1999)
 Contract: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu

TITLE

JOURNAL

COMMENT

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 Insert length: 1169 Std Error: 0.00
 High quality sequence stop: 386.

FEATURES

Source

1..452
 Location/Qualifiers
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1028-5081"
 /clone_1lb="Gm-c1028"
 /tissue_type="roots of 'Supernod' plants"
 /lab_host="DH10B"
 /note="Vector: pBluescript II Xr; Site 1: EcoRI; Site 2: XhoI; The mRNA was isolated from roots of Glycine max 'Supernod' plants generously donated by Dr. Gary Stacey. The seedlings were inoculated with Bradyrhizobium japonicus, strain USDA110 prior to harvest. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An 'anchor' nucleotide (V=A,C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGAGACTGCTCGAC(T)18V] to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA polymerase, the cDNA ends were filled in with subsequently phosphorylated. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GIBCOBRL life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratagene's pBluescript II Xr Predigested vector (pBluescript II SK(+)) that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=25). This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

BASE COUNT 134 a 111 c 90 g 117 t
 ORIGIN

Query Match 59.8%; Score 158; DB 24; Length 452;
 Best Local Similarity 99.4%; Pred. No. 1.2e-38;
 Matches 169; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 96 cgcctcactcaacttcttcacatcgctttaaatacaaaactaatcgtcgt 155
 Db 6 cgcctcactcaacttcttcacatcgctttaaatacaaaactaatcgtcgt 65

QY	156	tccttcgcattccattcaactgcctgcctccctcttaaaccccttaagatccaatgtcccaa	215
Db	66	TCCCTCCGATTCATTCCTACTGCTGCTCCCTCTTAAACCTCTTACGATCAATGCTCCAA	125
QY	216	acagatgggaatggaagtcctctgcaagaagagaca-tgcttcattgctctta	264
Db	126	ACAGATGGGATGGAAGTCTCTGCAAGAGGACACTGCTTCATGATCTTTA	175
RESULT	13		
LOCUS	BE347314		
DEFINITION	BE347314	690 bp	mRNA
VERSION	sp37e10.y1	Gm-cl043 glycine max cDNA clone	EST
KEYWORDS	Gm-cl043-451	5' similar to TR:004894	004894 TRANSALDOLASE ; mRNA
ORGANISM	BE347314		
SOURCE	BE347314.1	GI:9259167	
REFERENCE	soybean.		
AUTHORS	Glycine max		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
	Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids I;		
	Fabales; Fabaceae; Papilionoideae; Glycine.		
	1 (bases 1 to 690)		
	Shoemaker,R., Kaim,P., Vodka,L., Erpelding,J., Coryell,V., Khana		
	,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,		
	Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers		
	,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk		
	,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann		
	,R., Materson,R. and Wilson,R.		
TITLE	Public Soybean EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Shoemaker R/Public Soybean EST Project		
	Public Soybean EST Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@wustl.edu		
	This clone is available through: Genome Systems, Inc. 4633 World		
	Callway Circle St. Louis, Missouri 63134 For further information		
	call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)		
	427-3324 or contact: clones@genomesystems.com or		
	inf@genomesystems.com web site: www.genomesystems.com		
	High quality sequence stop: 485.		
FEATURES	Location/Qualifiers		
source	1..690		
	/organism="Glycine max"		
	/db_xref="taxon:3847"		
	/clone="GENOME SYSTEMS CLONE ID: Gm-cl043-451"		
	/clone_lib="Gm-cl043"		
	/tissue_type="Hypocotyl and Plumule, germinating seeds"		
	/lab_host="DH10B"		
	/note="Vector: p7T3Pac (Pharmacia); Site:1: EcoRI;		
	Site:2: NotI; This cDNA library was constructed from mRNA		
	isolated from hypocotyl and plumule tissues of seeds		
	germinated for three days of the cultivar Williams.		
	complementary DNA was synthesized from mRNA using a primer		
	consisting of a poly(dT) sequence with a NotI restriction		
	site. EcoRI adapters were ligated to the blunt-ended cDNA		
	fragments followed by digestion with EcoRI and NotI. The		
	cDNA fragments were directionally cloned into the		
	EcoRI-NotI restriction site of the p7T3-Pac vector. The		
	ligated cDNA fragments were transformed into DH10B host		
	cells (Gibco BRL). This library was constructed by Dr.		
	Randy Shoemaker."		
BASE COUNT	185 a	168 c	149 g
ORIGIN			187 t
			1 others
Query Match	55.8%	Score 147.4;	DB 35; Length 690;
Best Local Similarity	90.8%;	Pred. No.2.7e-35;	
*Matches 157; Conservative	0;	Mismatches 16;	Indels 0; Gaps 0

QY	27	tcgattccagcttcaaacgccaatcaactgtcttccttatcaagaagcgcttcgccc	86
Db	14	ttcatattcaaacctcttcgaacgcattctccactctctcttataagaaaggcttcgcccc	73
QY	87	cgagattctcgcttccctcaaccttcaaacctctcttcacatcgcttttaatacacaataaac	146
Db	74	agagattctcgcttccctctctctctcaaatcccttcttcgaacgctattatgacaaattaac	133
QY	147	ctatccgcttccctccgatcaatcaactaactgcgtccctctttaaactccctctacg	139
Db	134	ctctccgcttccctccgatcaatcaactaactgcgtccctctttaaactccctctacg	186
RESULT	14		
LOCUS	BE440344		
DEFINITION	sp44d01.y1 Gm-cl043 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl043-1082 5' similar to TR:004894 004894 TRANSLADOLASE ;, mRNA sequence.		
ACCESSION	BE440344		
VERSION	BE440344.1	GI:9439831	
KEYWORDS	EST.		
ORGANISM	soybean.		
SOURCE	glycine max		
REFERENCE	Shoemaker, R., Kelm, P., Vodka, L., Erpelting, J., Corryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and Wilson, R.		
AUTHORS	Public Soybean EST Project		
TITLE	Unpublished (1993)		
JOURNAL	Contact: Shoemaker R/Public Soybean EST Project		
COMMENT	Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com High quality sequence stop: 436.		
FEATURES	Location/Qualifiers		
SOURCE	1. 461 /organism="Glycine max" /db_xref="taxon:3847" /clone="GENOME SYSTEMS CLONE ID: Gm-cl043-1082" /clone_1lb="Gm-cl043" /clone_type="hypocotyl and plumule, germinating seeds" /lab_hosts="DH10B" /note="Vector: pRT73Pac (Pharmacia); Site_1: EcoRI; Site_2: NotI; This cDNA library was constructed from mRNA isolated from hypocotyl and plumule tissues of seeds germinated for three days of the cultivar Williams. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a NotI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by digestion with EcoRI and NotI. The cDNA fragments were directionally cloned into the EcoRI-NotI restriction site of the pRT73-Pac vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker."		
BASE COUNT	134 a 104 c 94 g 129 t		
ORIGIN			

```

BASE COUNT      111 a      77 c      56 g      78 t
ORIGIN

```

were directionally cloned into the EcoRI-XhoI restriction site of the plasmid vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker.^a

Query Match	36.7%;	Score 96.8;	DB 34;	Length 322;
Best Local Similarity	84.0%;	Pred No	1 1e-19.	

36.7%;	Score 96.8;	DB 34;	Length 322;
ity 84.0%;	Pred. No. 1.1e-19;		
servative 0;	Mismatches 22;	Indels 1;	Gaps 1;

QY 182 tcttaaacctccttacgqatcaaatgctcccaaacagatagaaatgaaatcctgaaa 341

61 TCTTAATACCTCCTTACGAAATCAATTAATCTCCCAACAGATATAAAGATTAAGTGTGCAAA 120

[illegible]

55 121 CAGGACAAATACATTCATGATCTTA 144

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

search completed: November 11, 2000, 18:47:51
Job time: 2076 sec

```

